Super - Resolution of Multi - Pixel and Sub - Pixel Images for the SDI

FINAL REPORT

Contract No. N00014 - 90 - C - 0173

Report Number 93 - 2881 - 06

June 08, 1993

Prepared For:

Office of Naval Research
Department of the Navy
800 N. Quincy Street
Arlington, VA 22217-5000

Prepared By:

TITAN Spectron
1582 Parkway Loop, Suite B
Tustin, CA 92680-6505

TABLE OF CONTENTS

		Page
ACKNOWLEDGEMENTS		
1.0	INTRODUCTION	4
2.0	THE SCENARIO AND THE PROBLEM	5
2.1	The Defensive Task	5
2.2	Typical Diffraction-Limited Image Properties	5
2.3	Deficiencies in the Sensor System	8
3.0	THE SOLUTION	9
3.1	Mathematical Background	9
3.2	Image Restoration Algorithms	11
3.3	Computational Aspects	24
3.3.1	Reducing Storage Space	24
3.3.2	Reducing Computational Complexity	25
4.0	WIDER APPLICATIONS	27
5.0	CONCLUSIONS AND RECOMMENDATIONS	28
6.0	REFERENCES	30
APPENDIX A	- REGULARIZED IMAGE RECONSTRUCTION	A-1
APPENDIX B	- ILLUSTRATIVE IMAGE-ENHANCEMENT PROGRAMS	B-1
APPENDIX C	- CONTRACT PUBLICATIONS AND PRESENTATIONS	C-1

ACKNOWLEDGEMENTS

The contributions made by Mr. Harper J. Whitehouse, Dr. Richard P. Bocker and Dr. Jeffrey C. Allen, of the Naval Command, Control & Ocean Surveillance Center, San Diego, to the program of research described herein are gratefully acknowledged. Dr. Keith Bromley of the Office of Naval Research administered the program and provided guidance and much helpful advice.

1.0 INTRODUCTION

The recent profound shift in the global balance of power in favor of the United States of America has had major repercussions on Strategic Defense Initiative (SDI) planning. In particular, the focus has shifted from the provision of protection for the United States against a massive raid, involving possibly thousands of reentry vehicles, to defense against a much more limited attack which could now, however, be launched from any part of the world. Additionally, the United States is seeking to protect its forces and allies overseas, and in the task of missile detection and tracking, allowance must now be made for trajectories which can begin and end in almost any inhabited area of the globe. Thus SDI demands on surveillance technology have been significantly expanded [1].

Space-based imaging systems will play a vital role in the surveillance task. In the detection, discrimination and subsequent tracking of hostile objects, image quality in particular will be of crucial importance. Considerations of size, weight, and cost will, however, impose strict limits on the unaided performance of the optical hardware. In addition, the sensor's operational capabilities may be impaired by system aberrations and degradations which may be inherent (as in the case of the Hubble telescope), induced by the stresses of launch or thermal fluctuations in orbit, or simply the result of aging processes in the space environment. If these defects can be properly characterized, however, newly developed algorithms can be used to compensate for them and thus restore the image. The resolving power of the system can also be extended in this way.

In this report a description is first given of a typical scenario. The potential imaging problems are then examined, the mathematical background is discussed, and the innovative algorithms which have been developed for correcting and enhancing the performance of the imaging sensor are described. Some results are included from current simulations based on the parameters of the typical scenario. The complexity involved in the calculations is assessed, together with the associated storage requirements. Attention is drawn to wider applications of the techniques developed and knowledge gained in the course of this work, and finally recommendations are made concerning the design attributes of an orbiting sensor network appropriate to the SDIO mission.

2.0 THE SCENARIO AND THE PROBLEM

2.1 The Defensive Task

Orbiting rings of small, lightweight surveillance satellites (the so-called 'brilliant eyes') provide the surveillance function from high earth orbit. They communicate among themselves, or via a command and control system to produce a multi-sensor map of the target area. The number, nature and trajectories of objects which are deemed to be potential threats, in the atmosphere or in space, are identified. Compact guided missiles (the so-called 'brilliant pebbles') are on stand-by and are dispatched by the command and control system to intercept specific objects among the group of threats. The missiles destroy or damage their targets through high-speed impact.

The optical requirements for the two groups of satellites (the 'eyes' and the 'pebbles') differ as a consequence of their intended functions. A pebble missile will be initially directed towards an intended target, which may be moving at a very high relative speed, and must be capable of acquiring and locking-on to it as rapidly as possible. Moreover, the missile must carry out these operations in a potentially very cluttered environment, and be capable of making critical final corrections to ensure success. Thus simple quadrant detectors and centroid techniques would not be appropriate except very near the target. The complexity of the pebble's sensor system can be greatly reduced by designing it to receive initial orientation instructions from the control center, then to achieve lock-on and make course corrections autonomously. The eyes, on the other hand, must be capable of surveying a wide field of view and of achieving high-resolution imaging to provide the information for detection, identification and tracking. The computational techniques discussed in this document are designed to enhance images of the type obtained by an eye, and would be applicable also to a pebble's optical system, if on-board resources permitted.

2.2 Typical Diffraction-Limited Image Properties

On the basis of certain assumptions about the characteristics of the optical system and of the target, we can make predictions of its performance. We shall estimate system resolution capabilities, fields of view and available image acquisition time for a moving target, and photon flux into the image.

We make the following assumptions for system parameters:

Wavelength	λ	=	10 μm	
Aperture	d	=	250 mm	
Focal length	f	=	1250 mm	(F/5 system)
Pixel width	p	=	25 μm	
Range to target	z	=	2000 km	

The range quoted is a maximum for an orbital height of 700 km.

Then:

Rayleigh resolution length at image plane		=	$1.22 \lambda f/d$	=	61 μm
Resolvable separation at range z		=	$1.22 \lambda z/d$	=	98 m
Airy disc diameter		=	$2.44 \lambda f/d$	=	122 μm
Pixels across Airy disc	=	$2.44 \lambda f/dp$		≈	5
Pixel angular field of view (FOV)		=	p/f	=	20μrad
FOV at 2000 km		=	40 m		

For a 1000-element linear staring array:

Total angular FOV	=	20 mrad x 20 μrad
Linear FOV at 2000 km	=	40 km x 40 m

In order to calculate the maximum time available for acquisition of a single image, we need an estimate of the maximum velocity of the object across the field of view. For re-entrant ballistic trajectories, we shall assume that this is of the order of the orbital speed for a low earth orbit, and use a value of 7 km s^{-1}. Then the image of a target at 2000 km range crosses one pixel, having a 40 m FOV, in approximately 6 ms. and the resolvable distance of 98 m in 14 ms. A reasonable image acquisition time might be half of this; i.e., 7 ms.

We now calculate the photon flux at the detector for a typical target.

Planck's formula for the power radiated per unit area, ΔI_ν, within the frequency interval $\Delta\nu$, by a black body at temperature T is [2]

$$\Delta I_\upsilon = \frac{2\pi h}{c^2} \cdot \frac{\upsilon^3 \Delta\nu}{\{\exp(h\upsilon / kT)\} - 1}$$

where h is Planck's constant (6.6×10^{-34} Js), k is Boltzman's constant (1.38×10^{-23} JK^{-1}) and c is the velocity of light (3×10^8 ms^{-1}).

For $\lambda = 10$ μm, $\upsilon = c/\lambda = 3 \times 10^{13}$ s^{-1}.

Suppose the body is radiating at 300 K and that the radiation is detected over a bandwidth of 9-11 μm. Then the corresponding frequency increment, $\Delta\upsilon$, is 0.6×10^{13} s^{-1} and from Planck's formula, the power per unit area ΔI_υ is approximately 63 W m^{-2}.

Suppose further the target is radiating into 4π steradians from an area of 1 m^2. Then the radiated power into the 250 mm aperture is approximately 6×10^{-14} W. The energy of a photon at 10 μm is 2×10^{-20} J. Hence the photon flux through the aperture is 3×10^6 s^{-1}. For an acquisition time of 7 ms, the number of photons contributing to the image will be 21×10^3. If the target is essentially a point source, these photons are distributed over the point spread function of the system. About 86% of this energy goes into the central lobe, the Airy disc, which covers about 25 pixels. Hence the average energy per pixel in the Airy disc is about 720 photons. The center pixel would in fact contain about 3800 photons.

Poisson statistics govern the energy distribution in the image around its expected value, and the signal-to-noise ratio (SNR) is given by the square root of the mean signal value in photons. For the average pixel energy of 720 photons, the corresponding SNR is 27; for the center pixel its value is about 61.

2.3 Deficiencies in the Sensor System

In the absence of any other degrading effects, the performance of an optical system is restricted ultimately by the effects of diffraction. The finite extent of the exit pupil imposes a fundamental upper limit on the system's spatial-frequency response. It is unlikely, however, that the image quality of an operational space-based system will approach the theoretical limit very closely at any time. It is possible that the design or construction will be flawed, as in the case of the Hubble telescope, through defective manufacture, assembly or quality assurance procedures. In addition, it is highly likely that the sensor will become degraded during the mission lifetime; faults may be induced by the stresses of launch and deployment, and aging of the equipment (caused, for example, by radiation or thermal effects) will almost certainly compromise component performance. The detector itself will impose limitations; for example, where a CCD array is used, information is lost in the inter-pixel areas, and image energy is integrated over the active area of each pixel. Other degrading factors will include defective pixels, noise in the CCD array and electronic subsystems, and the possibly obtrusive contributions of the earth-space background. Success in detection, identification and subsequent tracking will depend critically on the levels of noise and clutter in the images, and robustness is of fundamental importance in any image-processing scheme. The methods of image restoration considered here were originally aimed at achieving performance beyond the conventional diffraction limit [3, 4], but are in fact capable of compensating simultaneously or separately for aberrations induced by the optical components and for the limitations of the detector. They were designed to be robust and also possess valuable noise-suppression properties.

3.0 THE SOLUTION

3.1 Mathematical Background

The initial assumption is made that the effect of the optics can be described as a possibly time- and shift-variant blurring of the image due to diffraction and aberrations. Thus, in general, the point spread function will change across the sensor field of view. It can be assumed, however, that at any given time the point spread function is effectively constant, i.e., shift-invariant, over some localized area, and then undergoes a discrete change into neighboring areas. This is the case with the Hubble telescope. It will also be assumed that the set of point spread functions is known or can be measured. Under these circumstances the point spread function is said to be locally shift-invariant, and the image is created by the summation across the entire field of view of the set of localized point spread functions convolved with objects in the corresponding parts of the field. Mathematically, a convolution can be written as a Fredholm integral equation of the first kind. The solution of this class of equation is known to be ill-posed and numerically difficult [5, 6]. Additionally, it is anticipated that the image will be spatially sampled by a solid-state sensor which introduces spatial integration, discretization and associated noise processes. Thus, the Fredholm integral sum representing the continuous image can be rewritten as a matrix equation, to which modern signal processing techniques derived from advanced linear algebra can be applied. In general, the presence of the sensor noise takes the measured image out of the span of the columns of the kernel matrix, which is typically highly ill-conditioned; additional techniques derived from regularization theory are required to restore stability to the reconstruction. By introducing a suitable error criterion, images can be constructed which are, in terms of the chosen criterion, closer to the undistorted geometrical image of the object than the detected image data.

The point spread function of a well designed optical system is normally approximately invariant over large segments of the image plane. Note that imaging performance (and therefore the point spread function) commonly changes over relatively wide fields of view; for example, the resolution of typical photographic lenses is markedly less near the edge of the field than near the center. However, over some limited region the point spread function will be essentially shift invariant, although this is not mathematically necessary, and one may speak of an effective point spread function over this region with a transfer-function band-limit determined by the exit pupil of the system. System aberrations and pixel integration serve to reduce the system response within this passband.

To the extent permitted by the noise in the image, these in-band effects are relatively straightforward to remove for shift invariant systems: a typical approach would be to employ a Wiener-type inverse filter. However, detector pixellation and the finite aperture of any system set resolution limits not easily overcome, and a method for achieving spectral extrapolation has to be devised.

The spatial spectrum of the object is the Fourier transform of its amplitude, in the coherent case, or its intensity, in the incoherent case. If the object is known to be of finite extent, its Fourier transform is an analytic function, and the out-of-band part of the spectrum can in principle be fully recovered by analytic continuation [7] of the image spectrum after removal of any in-band distortion. The inverse Fourier transform of this extended spectrum would then yield a perfect image of the original object. Equivalently, one could attempt to solve directly the equation describing the imaging process. This, however, involves the inversion of an ill-conditioned matrix, and the restoration process is intrinsically unstable, even small amounts of noise rendering the results meaningless. These difficulties may be surmounted by applying the methods of regularization theory [8], developed to deal with ill-posed problems of this type, and the procedures which have evolved in the course of this SDIO/ISTO program are based on constrained least-squares methods in which a regularization parameter plays an essential role. Stability in the restored image, which is computed via the regularized pseudoinverse of the imaging matrix (see Appendix A), is controlled by this parameter. Its optimal value depends on the signal-to-noise ratio in the data. If it is not possible to select this optimum value from prior knowledge of the system characteristics and the object scene, several techniques are available for its estimation from the data themselves [9, 10]. Faster computation of the estimate is possible if the singular value decomposition of the imaging matrix is already available.

An efficient algorithm should exploit known properties of the target as much as possible. In the midcourse phase the targets of interest radiate thermally at approximately 300 K and the images are real, nonnegative distributions. The re-entry vehicles are relatively small, presenting an area of, say, 0.25 m^2, and are spinning, whereas the bus is larger, with an area of several square meters. All targets are in ballistic trajectories. The extent of the targets is much smaller at most ranges than the resolution limit of the optics described in Section 2.2, so that they typically act as point sources.

Effective image restoration requires as good an estimate as possible of the imaging system characteristics, careful design of the algorithm, appropriate stabilization of the algorithm against noise, and high computational efficiency. The algorithms which have been developed, and these which are still under development, differ primarily in the extent to which the target characteristics can be identified and incorporated into the algorithm, the dominant noise mechanisms in the imaging process, and the degree to which the computational efficiency can be optimized.

The most straightforward approach to obtaining the necessary imaging system characteristics for image enhancement would be to perform pre-launch measurements on a test rig. However, allowance should be made for the possibility of launch misalignments and in-orbit changes occurring; see Section 2.3. This can be done in two ways. The first is to modify the algorithms to allow for uncertainty in the calibration data as well as noise in the image data; the algorithms then depend on 'total' least-squares methods [11]. The second is to test the point spread function periodically and update the system data. In this context it may be noted that, if the eyes are capable of active surveillance, it may be possible to use the light source in one to calibrate another.

3.2 Image Restoration Algorithms

Reconstruction algorithms can be divided into two major classes: those for which the object can be allowed to take both positive and negative values, and those for which non-negativity is an important constraint. Algorithms in which non-negativity is imposed on the reconstruction are inevitably iterative and require greater computational effort to perform the minimization. The lower computational burden of the non-iterative form of the algorithms is available if non-negativity does not have to be imposed; this may be the case, for example, for relatively bright objects which stand out clearly from their surroundings. For a group of small isolated objects, on the other hand, use of the non-negativity criterion can be a powerful means of improving object location. Thus the methods, which also provide smoothed numerically stable solutions to the matrix equation, are well-suited for SDI space-based surveillance applications. Algorithms have been constructed for a number of specific purposes, including the correction of optical aberrations, for achieving resolution beyond the diffraction limit of the optical system and for the recovery of apparently lost sub-pixel detail. The performance of some of these algorithms will be illustrated with simulated and laboratory images. (See also Appendix B.) Preliminary studies have also been made of image data acquired during space launches by the ISTO Experimental Facility at Cape Canaveral; further analysis awaits more accurate information on the system point spread function.

The mathematical formalism underlying these restoration algorithms is given in Appendix A. The algorithms all incorporate regularization to counteract the destabilizing effects of noise, and require an estimate of the object support (i.e., the spatial region within which the target is non-zero). In the case of the non-negative reconstruction algorithm, this estimate is refined from one iteration to the next.

(i) Direct reconstruction from the regularized pseudo inverse.

This algorithm is widely applicable. It can be applied to finely sampled, nearly perfect images or to highly aberrated shift-variant images in which pixellation effects are clearly evident. The computation is direct - i.e., non-iterative - and the resulting solution can exhibit both positive and negative values. The reconstruction is bounded by the estimate of the object support and high resolution can be achieved. This method can give particularly good results when the unresolved targets of interest lie within a limited dynamic range. Weighting can be incorporated to reflect the level of confidence in the data and varying degrees of smoothness can be enforced on the reconstruction.

(a) Unweighted reconstruction

This practical example will demonstrate the use of the regularized pseudoinverse for image reconstruction with the identity matrix as the constraint operator (see Appendix A). Three small incoherently-illuminated apertures were imaged onto a CCD array and the optical geometry arranged so that the images of these objects all lay within one half the Rayleigh distance of each other. It can be seen from Fig. 1 that the individual images are unresolved. A good estimate of the system point spread function is essential for effective restoration, and Fig. 2 shows the result of averaging 16 estimates of the psf, obtained with a point light source, followed by filtering in the Fourier transform plane to eliminate out-of-band noise. The imaging aperture was constructed with a pair of crossed slits; the psf was therefore a product of two sinc functions. Fig. 3 is the computed reconstruction of the target field. The objects are now clearly resolved. Finally, Fig. 4 shows the image obtained with a large-aperture lens in place of the crossed slits. Contour plots of Figs. 3 and 4 show that the reconstructed images are in their correct relative locations.

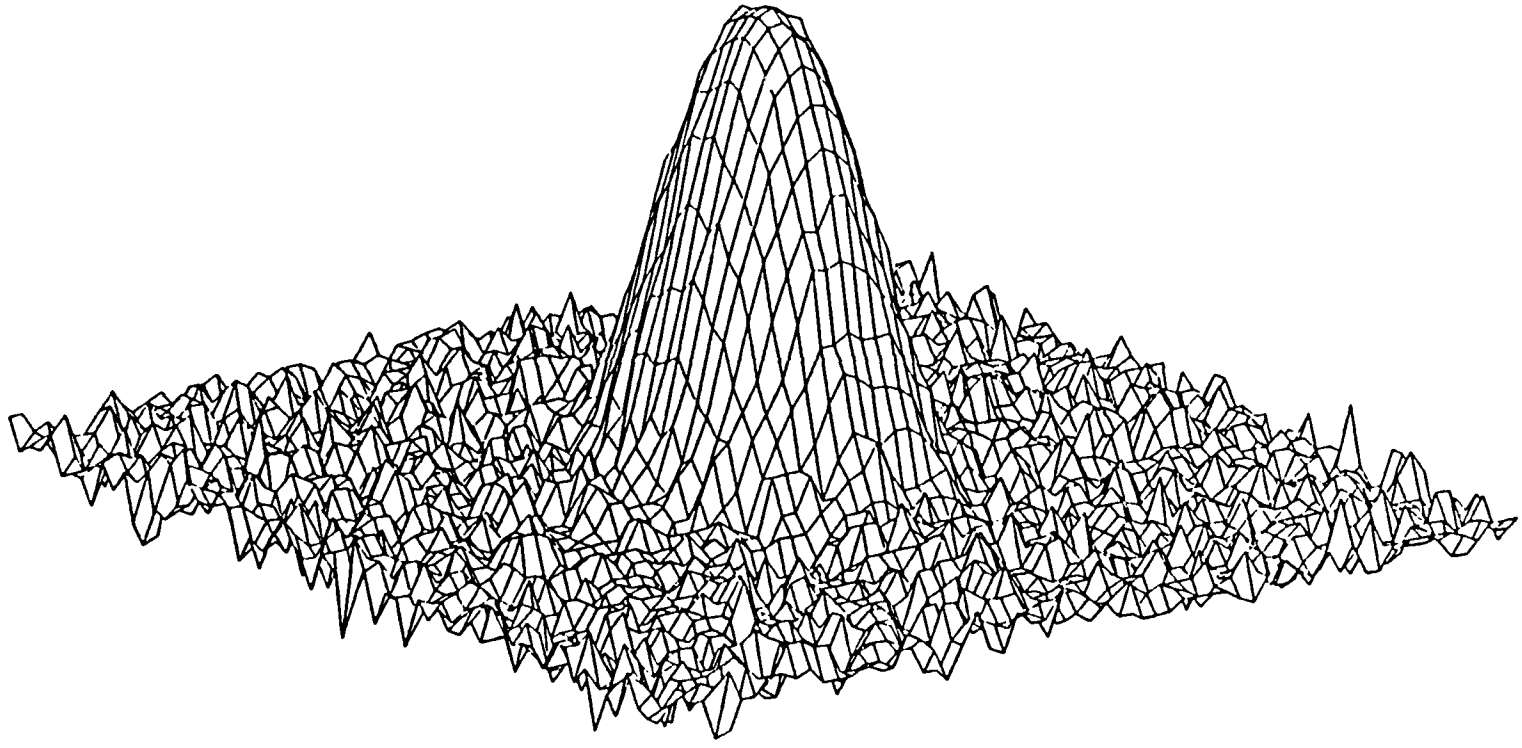


Figure 1. One Data Frame of Unresolved Object Field: Digitized to 7 Bits
(Sampling Approximately 4x Nyquist Rate)

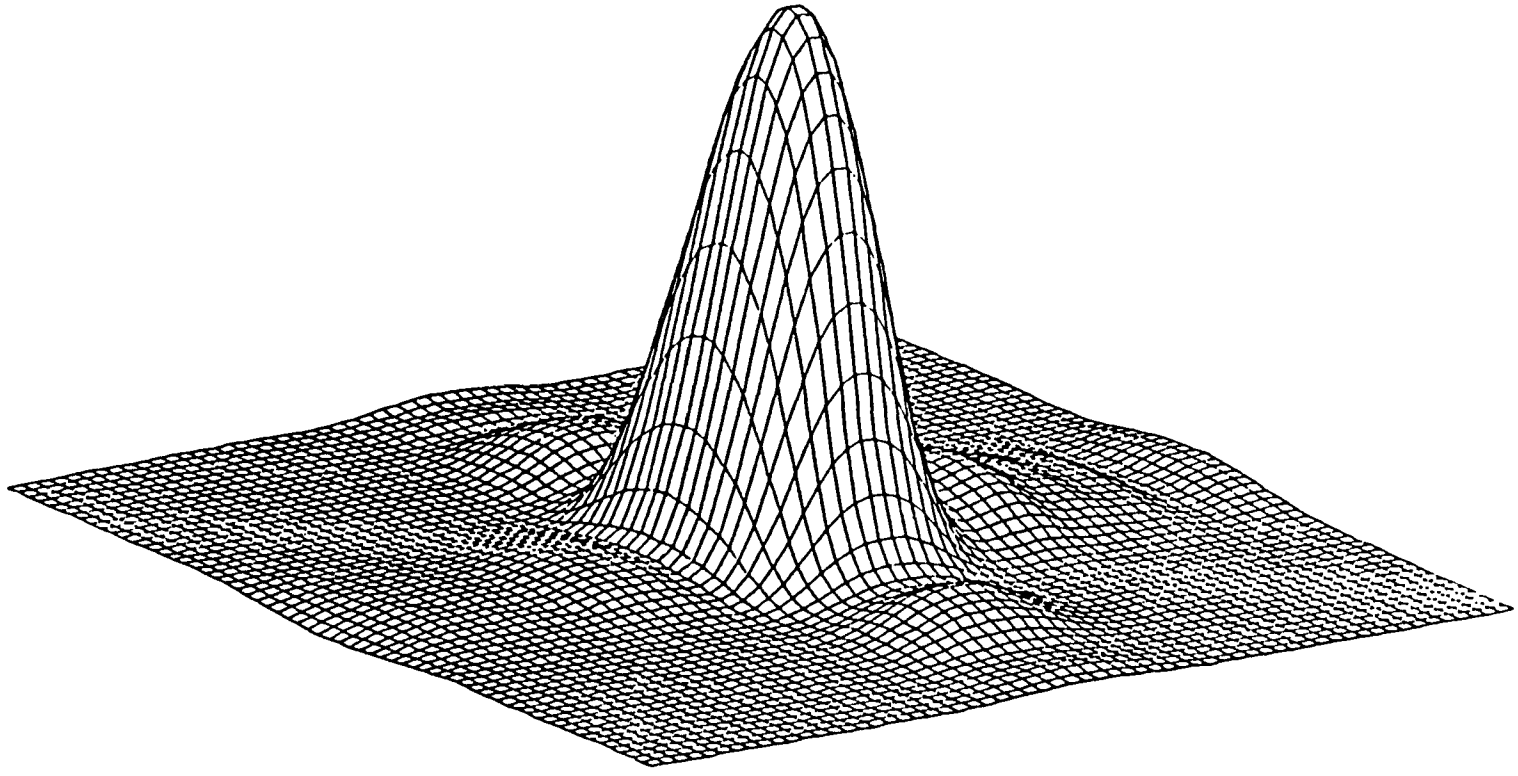


Figure 2. Filtered Averaged Point-Spread Function

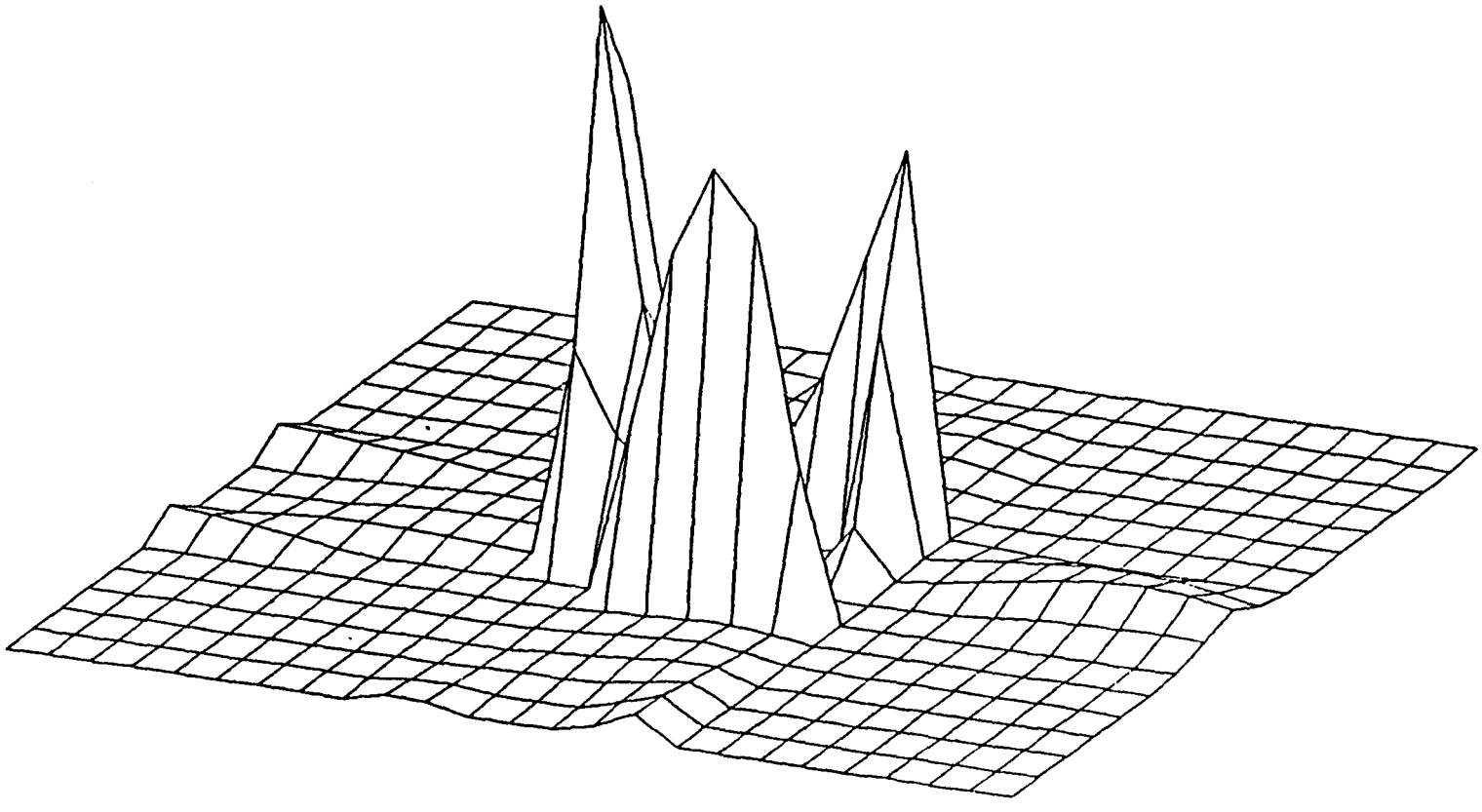


Figure 3. Reconstruction

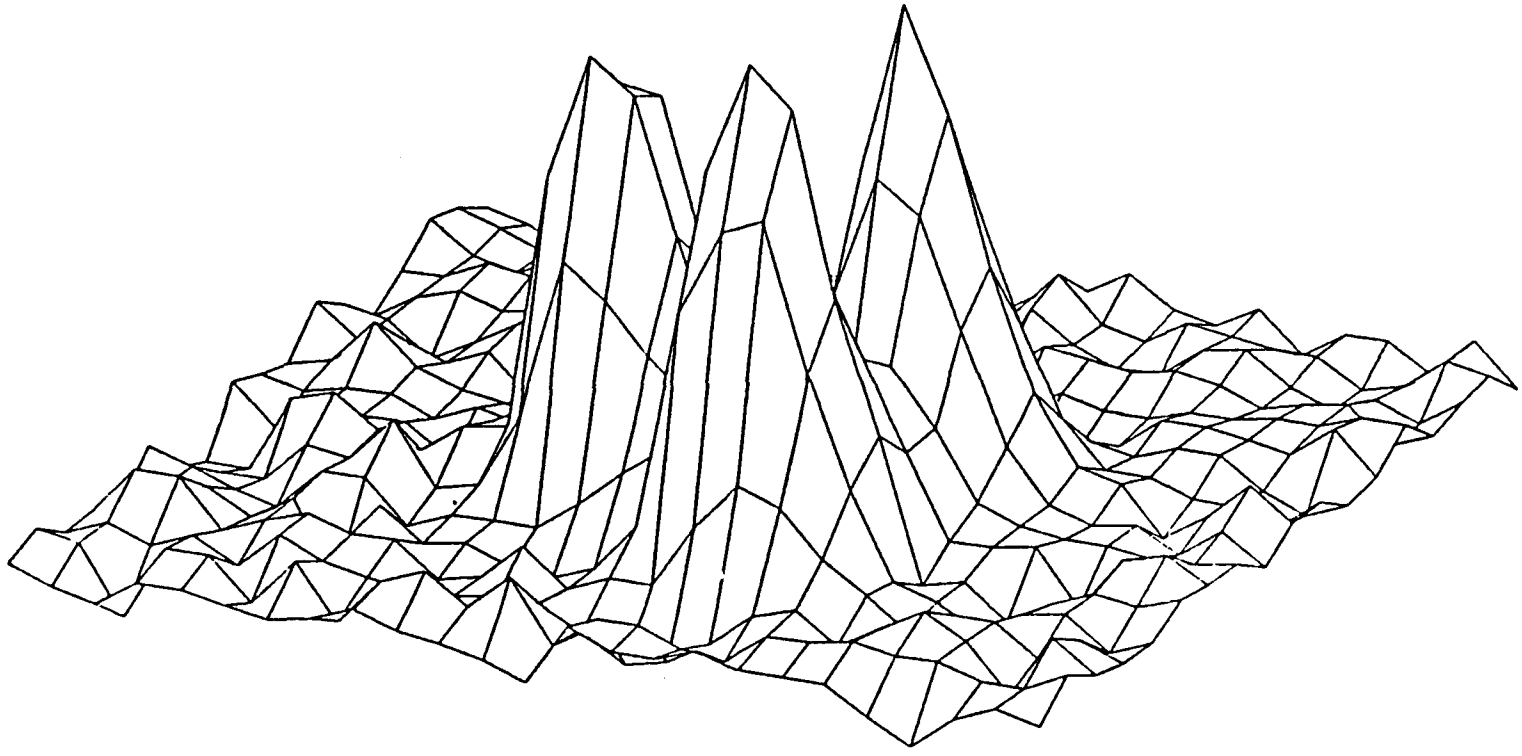


Figure 4. High resolution Image of Object

(b) Weighted reconstruction

Weighting can be introduced into the calculations in order, for example, to discriminate against bad pixels in the array, or to adjust the relative significance in the computation of bright and faint features in the target area. In the general case, the weight matrix is derived from a covariance matrix which expresses information on the correlations between the statistical errors in the image data values, as well as on their magnitude. If the pixel weightings are uncorrelated, the weight matrix is diagonal. This will be the case for an image in which fluctuations ('shot noise') due to the Poisson statistics of the detection process are the dominant noise mechanism.

(ii) Reconstruction of subpixel detail from combinations of defocused images

In addition to intrinsic noise, a CCD detector degrades the image information in two ways. First, there is necessarily a modest amount of averaging performed over the active area of the pixel, and second, the pixel structure imposes a spatial sampling rate. When dealing with CCD data sampled at the Nyquist rate of the imaging system the degradation caused by the detector is not great. Practically speaking, when imaging thermal (incoherent) radiation one requires approximately 5 samples across the point spread function core to achieve the Nyquist rate. When the optical design results in coarser sampling (for example, when the point spread function is matched to the pixel size) the detector itself severely limits the spatial resolution attainable. Noise from background sources or the detector elements themselves now limits the useful information to a relatively few pixels. As a consequence, although estimates can be made of the centroid of isolated point targets to subpixel precision, a single image cannot be made to yield detailed information about more complex object fields.

The discussion in Section 3.1 of the image restoration problem can be generalized and a method devised for extracting subpixel detail from a small number of independent images. These can be generated in a number of ways, such as by simultaneously combining defocused and/or laterally-shifted images of the same cluster of targets. One could also use time-series data, although the evolution of the image of a moving object over time may limit the resolution obtainable. Alternatively one could scan an image with a 1-D array so that high sampling rates are possible in the scan direction, and relatively low rates in the array direction. It is assumed that a pixel may contain images of more than one target, which poses a more difficult problem than that of locating the position of a single target with sub-pixel accuracy. A combination of correctly chosen

defocused images has been found to produce superior results to a combination of translated images [12]. The signal-to-noise requirements are more demanding than when pixel size is assumed to represent the fundamental limit on resolution.

(iii) Reconstruction with enforced non-negativity

Passive thermal imaging systems yield non-negative images, and this constraint may be important if the target area is relatively faint. Greater resolution, suppression of ringing and larger dynamic ranges are possible when a non-negativity constraint is imposed. It can be incorporated by iterating the computation a limited number of times; pixels with negative content can be eliminated after each iteration or discriminated against by weighting. In this manner one can adaptively obtain a very tight target support constraint with little a-priori information. Achieving tight support constraints gives the greatest degree of spectral extrapolation, and therefore resolution. Indeed, for small, localizable targets, the support estimate contains most of the desired targeting information.

Figs. 5-9 illustrate the performance of the non-negative algorithm in which negative pixels are eliminated at each iteration, in combination with the method for recovering sub-pixel detail described above. Fig. 5 shows the object field superimposed on nine pixels of a simulated CCD device, and Figs. 6 and 7 their images on the central seven by seven array at varying degrees of defocus. These four images are combined in sequence, and a composite matrix formed from the matrices representing the corresponding imaging operators. An iterative computation using the regularized pseudoinverse of this matrix, with non-negativity enforced at each step, was then carried out. The result for the 'noiseless' case is shown in Fig. 8. Note that the optimum value of the regularization parameter is non-zero even in the absence of added noise; computer roundoff error alone is sufficient to induce instability in the reconstruction. It can be seen that the individual sources have been correctly recovered in number and position. Fig. 9 shows the result obtained when Gaussian noise with a standard deviation of 5% of the peak pixel value was added to the data. One or two small artifacts have appeared and there is some blurring, but the individual targets are clearly identifiable.

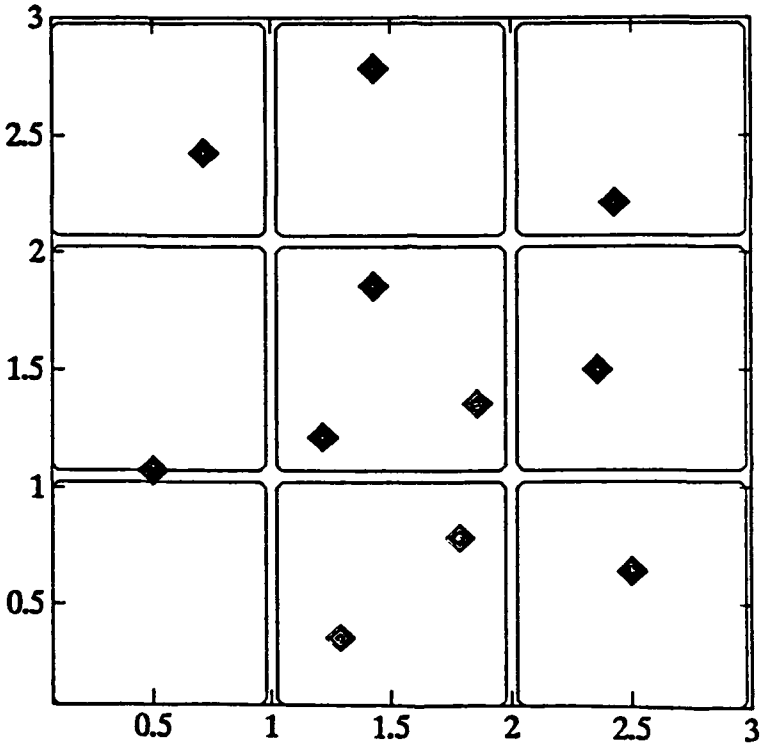


Figure 5. Object Field on Pixel Structure

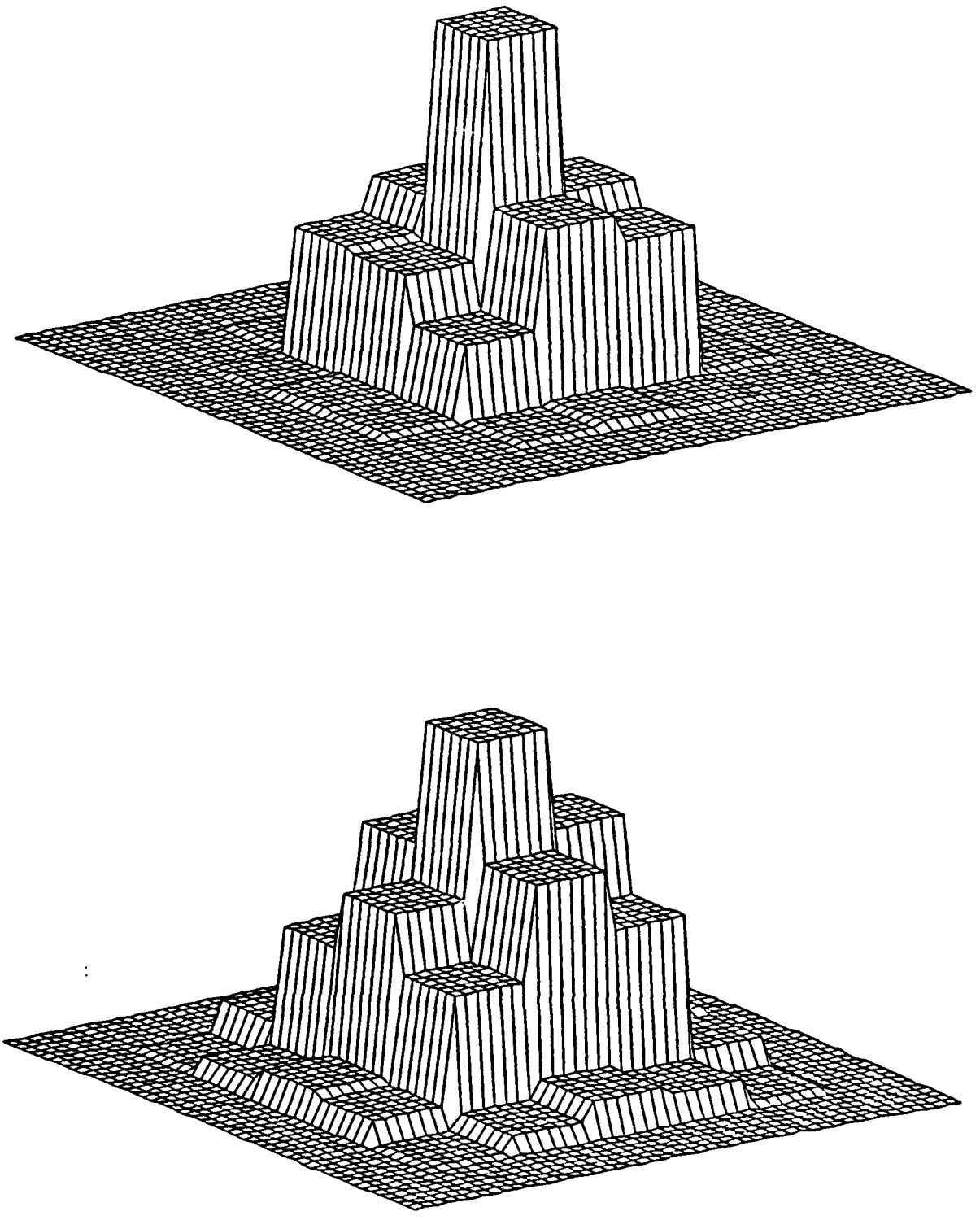


Figure 6. Images at 0 and 0.8 Waves Defocus

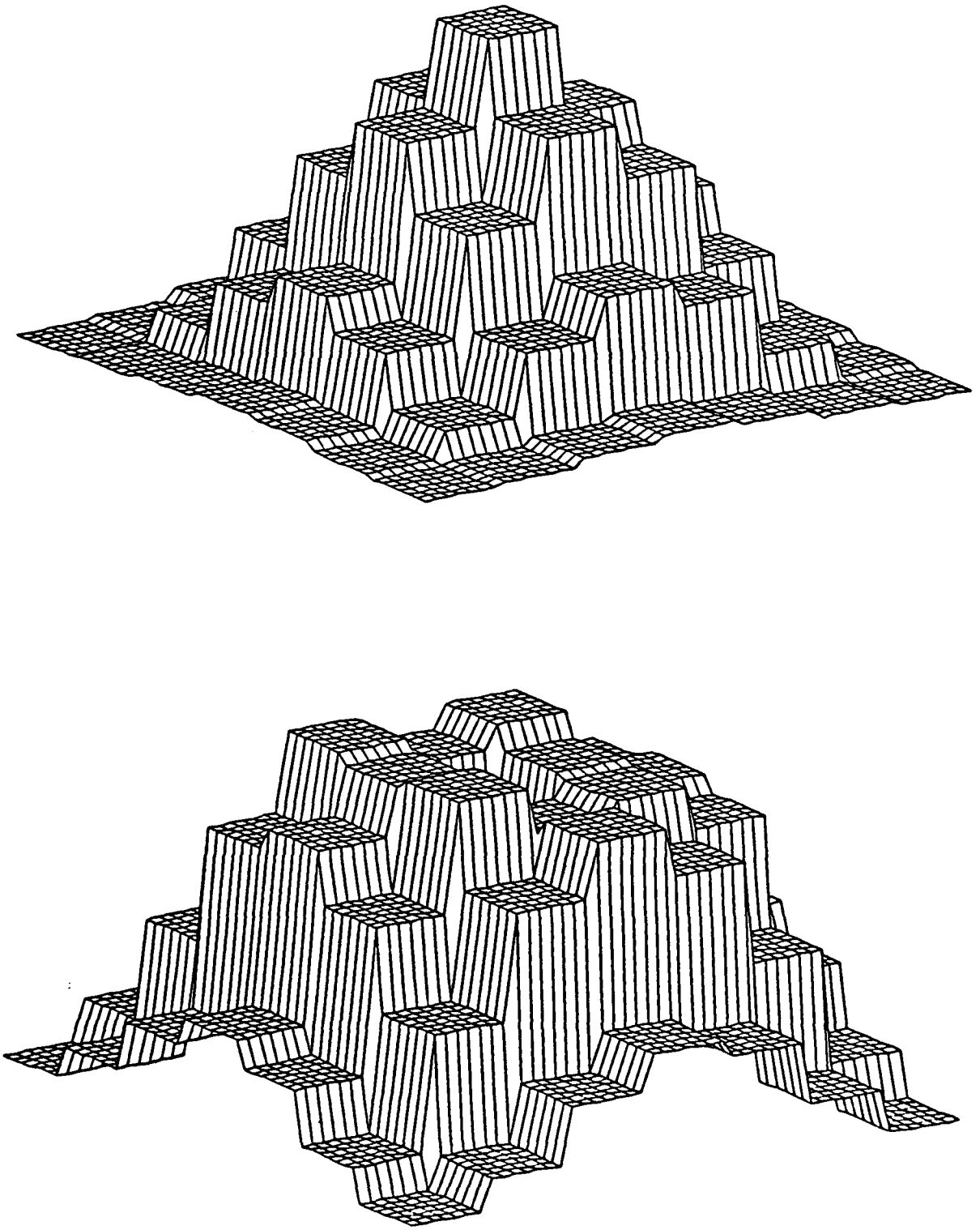


Figure 7. Images at 1.6 and 2.4 Waves Defocus

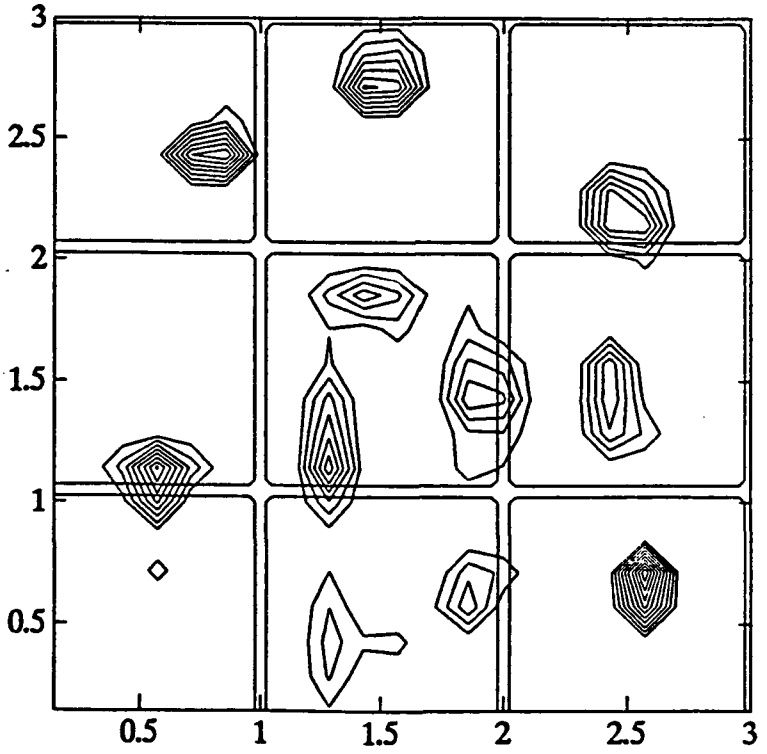


Figure 8. Reconstruction with 0% Added Noise; beta = 5e-5

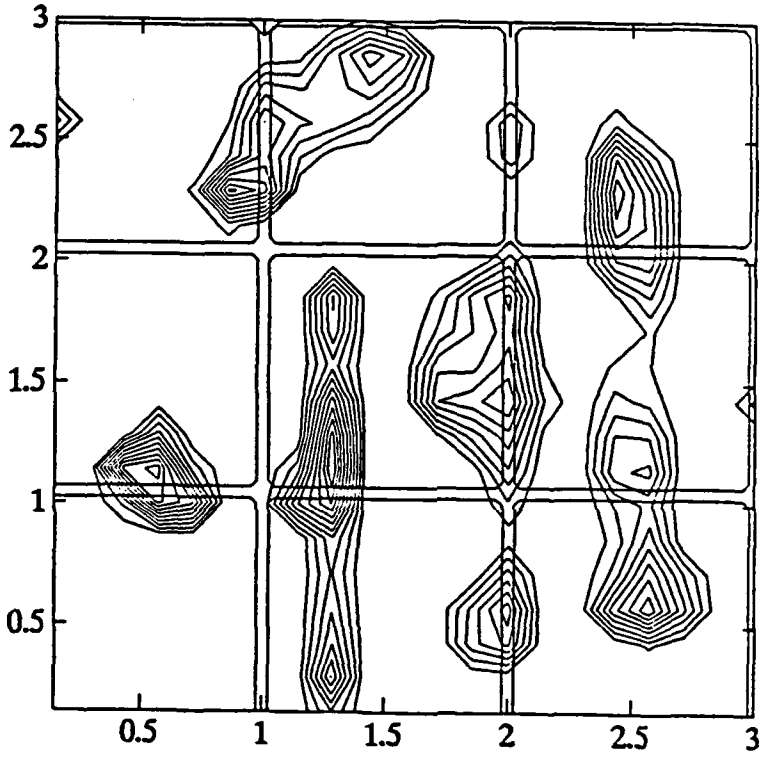


Figure 9. Reconstruction with 5% Added Noise; beta = 1e-3

3.3 Computational Aspects

An effective image restoration algorithm for SDI applications must not only exploit all available information about the target, it must also be well matched to the computational hardware and be carefully implemented to achieve optimum performance in near-real time. In implementing the algorithms described above, two difficulties become apparent as image size grows. First, the sheer number of computations can be very large, resulting in long computing times. Second, the matrix sizes in two dimensions quickly become overwhelmingly large. For example, a 64 x 64 image reconstructed into 64 x 64 elements has an imaging operator matrix of size $64^2 \times 64^2$, or 16.777×10^6 elements. For single precision operation (4 bytes/element), this results in 67 Mbytes of data (134 Mbytes with double precision). Handling such large data sets for modest images makes the high performance techniques described earlier challenging to implement even for earth-based applications, where extensive resources are available and time may not be a factor. Thus storage requirements and computational complexity are two aspects of algorithm implementation which must be addressed for all but the smallest images.

3.3.1 Reducing Storage Space

There are essentially three approaches to reducing space demands. The first, limited to shift invariant systems, is to use an iterative procedure based on the fast Fourier Transform (FFT). One can implement the regularized pseudoinverse reconstruction, including non-negativity, in this manner. The procedure is then a regularized form of the Gerchberg-Papoulis algorithm [13]. The data and image sets are the same size, and only FFT's are involved. Hence each iteration is fast and easily implemented; however, convergence is slow, and 1000 or more iterations may be required. The relative performance of this algorithm has not been evaluated.

The second method is to implement the regularized pseudoinverse exactly by partitioning the matrix equation into submatrices and using a sequential calculation based on smaller matrices. It has been shown that a real speed gain is possible for the pseudoinverse, although the expression for the regularized pseudoinverse has still to be explored. Smaller matrices are manipulated, and if the point spread function is shift invariant, some of the submatrices will be duplicated. Non-negativity probably cannot be implemented in this way.

The third method is to break the problem into smaller blocks and deal with each of the blocks independently [14]. The procedure is iterative, and potentially faster than solving the original problem as a unit. However, one must perform additional calculations to remove edge effects

from neighboring blocks. (The second method includes all cross-terms.) The ultimate efficiency of this method depends on the number of iterations required.

3.3.2 Reducing Computational Complexity

Considerable effort has been devoted, in collaboration with members of NOSC staff, to developing techniques for alleviating the potential computational burden associated with the algorithms discussed above [15, 16]. If the image restoration matrix has to be determined for each image, the number of operations involved in computing the restoration will be $O(N^6)$ for an $N \times N$ image. For even modest values of N, this number will be very large; for example, if $N = 32$, the number of operations is $O(10^9)$.

The entries in the imaging matrix A consist of discrete values of the system point spread function. These elements are determined by the locations of the samples in the image plane and the locations of the points at which the corrected values are to be determined in the reconstruction plane. The point spread function will in general be locally shift-invariant, but change gradually across the field of view, and it will be assumed that the instrument is fully characterized by a stored set of calibration data. If the point spread function is also time-varying, recalibration must be performed as often as necessary with some appropriate source; if they are suitably equipped, one of the 'eyes' could possibly fulfill this function. Thus, for a given target area, we can assume that the required point spread function is already available. Suppose that image size and reconstruction support (the region within which the target objects are estimated to lie), or a typical set of them, are known in advance. Then only a value for the regularization parameter is required, and it is quite likely that a small discrete set of these will suffice. (It may be noted here that truncating the singular values of A at the appropriate point can be shown to be equivalent to incorporating a regularization parameter in the calculation [17], and may offer some advantages in terms of parallel processing efficiency.) Then a set of restoration matrices can be precomputed and the processing of the image reduced to a single vector-matrix multiplication involving $O(N^4)$ operations. This can still be a daunting task; for a 100 x 100 image, $O(10^8)$ operations are required. We therefore look for special mathematical features in the structure of the problem which may be exploited to accelerate the computation further.

The imaging operator, from which A is derived, is a convolution for the shift-invariant system and has natural cyclic properties; in the case of equal numbers of samples and equal sampling intervals in image and reconstruction spaces, the matrix A is Toeplitz. Even if these conditions are not met, A may retain much Toeplitz-like structure. The concept of displacement rank [18]

may be used to exploit this property; the singular value decomposition of A then provides a means for constructing approximations to the image-restoration matrix (the regularized pseudo-inverse of A) in the form of relatively low-order sums of triangular Toeplitz matrix products. Fast multiplication of the image vector by the image-restoration matrix then becomes possible [19]. Further theoretical developments have also recently appeared in the open literature [20, 21].

For the two-dimensional image, the imaging matrix A can be written as a matrix consisting of Toeplitz sub-matrices arranged in Toeplitz order (block-Toeplitz with Toeplitz blocks). Then it can be shown that the restoration matrix R (the regularized pseudoinverse of A) is block-persymmetric with persymmetric blocks. Thus it retains marked cyclic features. On the basis of numerical experiments, it has been proposed that a circulant approximation could be made to R , which would vastly accelerate the final computation, since FFT techniques would now become available. Recent work [22] in this field provides supporting evidence that under certain conditions a series expansion may exist for R in which the first term is a circulant. The real significance of the proposal is, of course, that the essential information needed to carry out the reconstruction is contained in a single row or column of R . A further possibility now suggests itself, that there may be a more direct way to obtain this information than via the QR or singular-value decomposition of A . In the most recent phase of this program, the collaborative research between Spectron and NCCOSC has concentrated on this topic.

Displacement rank and related expansions are naturally suited to algorithm implementation on parallel processors, and the latter has been kept constantly in mind in the course of this program. The heavy computational demands associated with image reconstruction will be most effectively satisfied by a judicious combination of advanced techniques of matrix algebra and a carefully-matched parallel-processing architecture.

4.0 WIDER APPLICATIONS

The key to the applicability, for any given task, of the algorithms discussed in this document is that the "image" and the "object" are related by a linear transformation. Then, for a sampled data system, the discrete representation of the image (the sensor data) and the discrete approximation to the object will be related by a matrix. The matrix incorporates all the evolutionary processes undergone by the information-carrying field during propagation from the object and collection, detection and digitization by the sensor system. In the case of a conventional optical image, the field is electromagnetic, and the transformation consists primarily of Fourier transforms; the sensed data may be further modified, of course, by the characteristics of the detector and the associated circuitry. Nonlinear effects, such as Poisson noise in the optical case, can also be accommodated provided they are not so great that the assumption of a linear object-image relationship ceases to be a reasonable approximation.

Potential applications exist in many areas of remote sensing, including tomography, synthetic aperture radar, imaging radar, sonar, seismic prospecting, spectroscopy and system identification and control. In the imaging radar case, for example, tomographic reconstruction of the scattering function via the squared modulus of the transmitted signal's ambiguity function (the analog of the optical point spread function) would yield enhanced resolution [23]. In the case of synthetic aperture radar, the point spread function is separable and particularly efficient reconstructions of the scattering distribution are possible. It may also be noted that missing data segments, which are a common feature of inverse scattering problems, can be recovered with the techniques described here.

5.0 CONCLUSIONS AND RECOMMENDATIONS

As a result of the major geopolitical events which have recently occurred, the nature of the ballistic missile attack threat to the United States has undergone significant change. Many additional areas of the globe have become potential launch sites, and the defense of regions outside the United States is now declared public policy. The surveillance function is a fundamental component of defensive strategy, and a clear implication of these developments is that a comprehensive network of space-based observation platforms will be needed to meet the SDI requirements for detection, discrimination and subsequent tracking. In fulfilling the surveillance mission, the performance of the on-board optical sensors and the quality of the images they create will be of paramount importance.

Given adequate illumination, the performance of an optical system is limited ultimately by the physical laws governing the phenomenon of diffraction. Further limitations can arise from design errors, faulty construction, the stresses of launch and deployment or the effects of extended operation in the space environment. This report has been concerned with the exploitation of modern mathematical techniques, in combination with state-of-the-art computational technology, to compensate for such defects and deficiencies. The feasibility of recovering target information apparently irretrievably lost in the process of forming and capturing the image has been demonstrated with numerous simulations as well as experimentally-derived data. As described above, algorithms have been designed and developed for aberration correction, to achieve resolution beyond the diffraction limit and for other specific image-enhancing purposes which require only the provision of appropriate computing resources. Powerful techniques of linear algebra exist, including, for example, displaced-difference expansions and circulant approximations, which can be exploited to reduce the associated computational burdens. Further research is needed in this area into the ranges of validity of the approximations and their most efficient implementation.

The possible problems which may arise from the stresses of launch and deployment, from errors of manufacture which may not be discovered until the system is in orbit and operational, and from aging of the optical and electronic components, imply that provision for recalibration of the optical subsystem must be made. This can most easily be accomplished using a cooperative target equipped with a light source, which could be another eye, if these are active. Thus communications channels between the orbiting eyes will be needed. In addition, computation of the modified reconstruction matrix and subsequent updating of the on-board data would be most efficiently performed on a large ground-based computer, requiring a ground-orbit communication

link for downloading of the calibration data and uploading of the new matrix elements. State-of-the-art computer technology would be available for the near real-time correction of system deficiencies on a continuing basis, thereby considerably extending the useful lifetime of the orbiting sensor network.

It is therefore recommended that, in addition to continuation of the research into algorithm structure and implementations, provision be made for both adequate computing resources and appropriate sensor-to-sensor and ground-to-sensor communication links in the basic design of the surveillance sensor system, thus making possible both the correction of image deficiencies and the efficient updating, as necessary, of the computational data base. The benefits will include enhanced detection, discrimination and tracking capabilities and a longer-lived, more cost-effective defensive screen.

6.0 REFERENCES

[1] G. J. Dezenberg, "SDIO Ladar Technology Program Update", *Advanced Technology for Applications*, Proc. SPIE 1633, 1992.

[2] R. H. Kingston, *Detection of Optical and Infrared Radiation*, Springer-Verlag, New York, 1978.

[3] J. B. Abbiss, C. De Mol and H. S. Dhadwal, "Regularized Iterative and Non-Iterative Procedures for Object Restoration from Experimental Data," Optica Acta, Vol. 30, 1, 107-124, 1983.

[4] J. B. Abbiss and P. G. Earwicker, "Compact Operator Equations, Regularization and Super-Resolution", Proceedings, IMA Conference on Mathematics in Signal Processing, University of Bath, Bath, United Kingdom, 17-19 September 1985. (Published by Clarendon Press, Oxford, U.K., 1987.)

[5] Peter A. Jansson (Ed.), *Deconvolution with Applications in Spectroscopy*, Academic Press, New York, 1984.

[6] C. K. Rushforth, "Signal Restoration, Functional Analysis, and Fredholm Integral Equations of the First Kind", in: Henry Stark (Ed.), *Image Recovery: Theory and Application*, Academic Press, New York, 1987.

[7] M. A. Evgrafov, *Analytic Functions*, reprinted by Dover Publications, New York, 1978.

[8] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington DC, 1977.

[9] G. Wahba, "Practical Approximate Solutions to Linear Operator Equations When the Data are Noisy," SIAM J. Numer. Anal., Vol. 14, 4, 651-667, 1977.

[10] P. C. Hansen, "Analysis of Discrete Ill-Posed Problems by Means of the L-Curve," Argonne National Laboratory Preprint MCS-P157-0690, July 1990.

[11] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 2nd Edn., 1989.

[12] J. B. Abbiss and B. J. Brames, "Restoration of Sub-Pixel Detail using the Regularized Pseudo-Inverse of the Imaging Operator", *Advanced Signal Processing Algorithms, Architectures and Implementations II*, Franklin T. Luk, Editor, Proc. SPIE 1566, 365-375, 1991. (Reproduced in Appendix C.)

[13] R. W. Gerchberg, "Super-Resolution through Error Energy Reduction," Optica Acta, Vol. 21, 709-720, 1974.

[14] Don J. Lindler and Sara R. Heap, "Block Iterative Restoration of Astronomical Images with the NASA/Goddard Massively Parallel Processor," Proceedings, Second International Conference on Supercomputing, International Supercomputing Institute, 1987.

[15] J. M. Speiser, H. J. Whitehouse and J. C. Allen, "Fast Matrix-Vector Multiplication Using Displacement Rank Approximation via an SVD", Proceedings, Second International Workshop on SVD and Signal Processing, University of Rhode Island, Rhode Island, 25-27 June 1990.

[16] J. B. Abbiss, B. J. Brames and M. A. Fiddy, "Computational Aspects of Regularized Image Reconstruction". *Digital Image Synthesis and Inverse Optics*, Arthur F. Gmitro, Paul S. Idell and Ivan J. LaHaie, Editors, Proc. SPIE 1351, 85-96, 1990. (Reproduced in Appendix C.)

[17] P. C. Hansen, "Truncated Singular Value Decomposition Solutions to Discrete Ill-Posed Problems with Ill-Determined Numerical Rank," SIAM J. Sci. Stat. Comput., Vol. 11, No. 3, 503-518, May 1990.

[18] T. Kailath, S-Y Kung and M. Morf, "Displacement Ranks of Matrices and Linear Equations," J. Math. Anal. and Appl., Vol. 68, 395-407, 1979.

[19] J. B. Abbiss, J. C. Allen, R. P. Bocker and H. J. Whitehouse, "Fast Image Reconstruction based on the Regularized Pseudo-Inverse of the Imaging Operator", *Inverse Problems in Scattering and Imaging*, Michael A. Fiddy, Editor, Proc. SPIE 1767, 93-111, 1992. (Reproduced in Appendix C.)

[20] J. Chun and T. Kailath, "Divide-and-Conquer Solutions of Least-Squares Problems for Matrices with Displacement Structure," SIAM J. Matrix Anal. Appl., Vol. 12, 1, 128-145, 1991.

[21] T-K Ku and C. C. Jay Kuo, "Design and Analysis of Toeplitz Preconditioners", IEEE Trans. Signal Processing, Vol. 40, No. 1, pp. 129-141, January 1992.

[22] G. Ammar and P. Gader, "New Decompositions of the Inverse of a Toeplitz Matrix," in: A. A. Kaaschock, J. H. van Schuppen and A. C. M. Ran (Eds.), *Signal Processing, Scattering and Operator Theory, and Numerical Methods*, Birkhäuser, Boston, 1990.

[23] J. B. Abbiss, J. C. Allen, R. P. Bocker and H. J. Whitehouse, "Fast Regularized Deconvolution in Optics and Radar", Third IMA Conference on Mathematics in Signal Processing, University of Warwick, Warwick, UK, 15-17 December 1992. (Reproduced in Appendix C.)

APPENDIX A

REGULARIZED IMAGE RECONSTRUCTION

The task of image reconstruction belongs to the more general class of ill-posed inverse problems, which arise in many fields besides optics - for example in radar, geophysics, tomography or the extrapolation of band-limited signals. Suppose we are given the operator equation

$$Ax = y$$

where $A : X \to Y$. The problem of finding the solution x, given y, is said to be well-posed if for each y there exists a unique solution which depends continuously on the data; i.e., is stable under small changes in the data [G]. Otherwise, the problem is ill-posed. It should be emphasized that ill-posedness is inherent in the physical nature of the measurement process, not the mathematical modeling of it.

Consider a typical image reconstruction problem. We suppose some object f, an element of Hilbert space F, is imaged by a linear operator A into an element of another Hilbert space G, and that this image is perturbed by an unknown additive noise function which may arise in either the imaging or the observation process. The effect of the noise in practice will generally be to displace the image g out of the range of A - there will no longer be an element of F corresponding to g and the equation $Af = g$ will have no solution. Clearly the problem is ill-posed.

This difficulty of a lack of continuous dependence can be overcome by one of the various techniques of "regularization." Essentially, the ill-posed problem is replaced by a related well-posed one, chosen to be physically meaningful and to possess the necessary properties of convergence and stability. Thus we may change the concept of a solution, or the Hilbert spaces or their topologies, or the operator itself. The technique we shall use belongs to the last category.

We impose physically reasonable constraints on the permitted solutions. If ε is a measure in the norm sense of the noise in the image (norm in the appropriate Hilbert space is denoted by $\| \cdot \|_H$) and if C is some constraint operator with E a known bound, we shall require that all possible reconstructions f' satisfy

$$\|g - Af'\|_G \leq \varepsilon \text{ and } \|Cf'\|_F \leq E$$

C may be used, for example, to impose smoothness on the reconstruction or to weight the reconstruction support. If C is the identity operator, E is a bound on the norm of the reconstruction. We combine the constraints quadratically and minimize the functional

$$\|g - Af\|_G^2 + \beta \|Cf\|_F^2$$

where $\beta = \varepsilon^2 / E^2$. Note that smaller values of β are equivalent to demanding greater fidelity between the reconstruction and the data; greater values place more emphasis on the property of the reconstruction associated with C. In the present discussion we shall take $C = I$. The minimizer f_β can be expressed in either of the forms

$$f_\beta = (A^*A + \beta I)^{-1} A^* g$$

or

$$f_\beta = A^*(AA^* + \beta I)^{-1} g$$

where A^* is the operator adjoint to A. The inverses of the bracketed operators will always exist, since the eigenvalues of the symmetric operators A^*A and AA^* are non-negative.

The operator, the image and the reconstruction will consist in practice of finite arrays. We shall assume that the imaging operator has been recast (if necessary) as a two-dimensional matrix and that f and g are vectors. The finite-dimensional approximations to the Hilbert spaces F and G are K^N and K^M, N- and M-dimensional spaces of real or complex numbers. For matrices with complex entries, A^* becomes A^H and the solution is

$$f_\beta = Rg \ , \tag{1}$$

where R can take either of the forms

$$\left(A^H A + \beta I_N\right)^{-1} A^H g$$

or

$$A^H\left(AA^H + \beta I_M\right)^{-1} g \ .$$

Either of these matrix expressions is referred to as the regularized pseudoinverse of A. To facilitate analysis and gain insight into the nature of the ill-posed problem we make use of the method of singular value decomposition (SVD). (There is also a significant gain in accuracy in using the SVD rather than in forming A^HA directly.) We write for the SVD of A

$$A = U\Sigma V^H \tag{2}$$

where

$$U^HU = V^HV = VV^H = I_N$$

and

$$\Sigma = diag(\sigma_1 .... \sigma_N), \ \sigma_1 \geq 0.$$

U consists of the N othonormalized eigenvectors u_i associated with the N largest eigenvalues of AA^H, and V consists of the N orthonormalized eigenvectors v_i of A^HA. The σ_i, the non-negative square roots of the eigenvalues of A^HA, are the singular values of A.

Then, from equation (1),

$$f_\beta = V \ diag\left(..., \ \sigma_i / \left(\sigma_i^2 + \beta\right), ...\right) U^H g \tag{3}$$

If $\beta = 0$, we have

$$f_\beta = A^+ g$$

where A^+ denotes the Moore-Penrose generalized inverse (familiarly called the pseudoinverse) of A. Thus the matrix operator in equation (1) can accurately be described as a regularized pseudoinverse of A.

The regularizing role of β can clearly be seen in equation (3); for zero β, the elements of the diagonal matrix are simply the inverses of the non-zero singular values of A and can grow without bound as these tend toward zero.

Weighting can be introduced into both of the spaces F and G through the definition of the inner product, and can be used to predispose the shape of the reconstruction to some prior expectation, or to reflect statistical information about the noise. A modified form of equation (1), incorporating the weighting matrices, now governs the reconstruction, which with the aid of Cholesky factorization can again be analyzed by SVD [D].

APPENDIX B

ILLUSTRATIVE IMAGE-ENHANCEMENT PROGRAMS

A suite of operational programs based on selected image-enhancement algorithms which were developed under this contract can be found on the floppy disk included herewith. The programs are designed to run in the MATLAB environment on a Sun workstation. Their individual functions are summarized below. The reconstruction procedures are based on the assumption that the point-spread function (psf) is two-dimensional and non-separable.

A demonstration is available by typing >>startup2d. This routine generates and saves the psf and the singular-value decomposition of the psf for the demonstration procedure. For convenience, (sinc(x)*sinc(y)).^2 was used for the psf. This function is, in fact, separable, although the computation is performed as if it were non-separable. Note that 'infinities' are generated in attempting to calculate sin(0)/(0) ; these result in standard MATLAB warning messages, but have no effect on the reconstruction.

demo2d_4 is called at the end of the startup2d routine; after the first setup run, it can be called directly. The program then randomly constructs an object, forms its image, adds the specified amount of noise and calculates a value for the regularization parameter using the technique of weighted cross-validation. Finally the reconstruction is computed, using two different methods for estimating the object support. The results are displayed in two final graphics screens (note the <PAUSE> between them).

Reasonably accurate estimation of the object support is an important requirement if good reconstructions of an object are to be achieved. In this demonstration, the first method for support estimation uses a reconstruction space which is coextensive with the space over which the object was generated. This constraint is very weak - it would be somewhat better, for example, to use a smaller bound estimated from the image width and the psf width - and the resulting reconstruction is often not a great improvement over the original image. However, the second algorithm (nnrec_w2) generally performs extremely well. In estimating the reconstruction support, this algorithm uses the fact that the object should be non-negative. It calculates the reconstruction in 10 iterations (the succession of 2-d reconstructions mapped column-by-column to vectors is displayed during this process). The assumption is made after each step that any negative points must be outside the object support, and a new reconstruction is generated with the

modified support. This procedure is most powerful in the 2-d nonseparable case, and cannot be effectively applied to 2d-separable problems where the separability is exploited in the computations, since changing one pixel in one of the dimensions will in that case affect an entire row of the image.

The figures at the end of this Appendix are an example of the output of the program demo2d_4.

Program Function Summary

2-D Mapping Algorithms

mat2vec	Takes a matrix (such as an image) and maps it to a vector.
vec2mat	Takes a vector and maps it to a matrix.
mul	A convenient procedure for multiplying A and f, where A is the psf matrix and f is in the form of a 2-d image. mul maps f to a vector, performs the multiplication, and remaps A*f back into a 2-d image.

Point-Spread Function Definition

mkpsf2d_sinc	Forms sinc^2 psf matrix. This psf corresponds to a square unapodized aperture (a slit).
mkpsf2d_sinc12 :	Forms sinc^2 psf matrix with the sampling rate in image space equal to half that of object space. This psf corresponds to a square unapodized aperture (a slit).

Regularized Matrix Inversion

oper	Forms inv(A'A + beta*I)A' using MATLAB inv function.
operpen	Forms the pseudo-inverse of A from the svd of A.
opersvd	Forms regularized pseudo-inverse from svd of A.
operwf	Forms inverse of inv(A'A+betaW'W)A', where W is specified by the user, and weights the reconstruction norm.

Reconstruction Algorithms

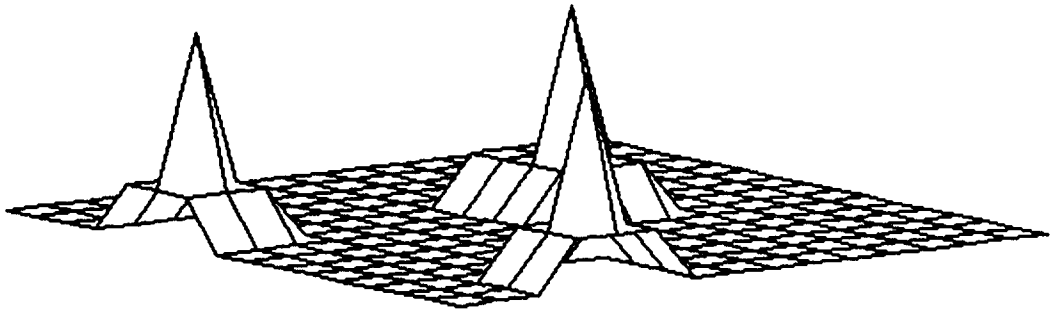
recld	Calculates the regularized pseudo-inverse Q and the reconstruction. Allows for the specification of a convex support constraint.
recld_g	Calculates a reconstruction using Gaussian elimination. Allows for the specification of a convex support constraint.
recld_s	Calculates a reconstruction using svd of psf. Allows for the specification of a convex support constraint.
nnrec_g1	Calculates non-negative reconstruction using Gaussian elimination. Effectively derives support constraint in the process. The support decreases from iteration to iteration, speeding the calculation, but there is a risk of eliminating pixels within the true support.
nnrec_g2	Essentially the same as nnrec_g1, except that the regularization parameter decreases with each iteration.
nnrec_w1	Calculates a non-negative reconstruction using Gaussian elimination. Uses inner product weighting to effectively derive a support constraint. Differs from nnrec_g1 in that computational time per iteration is constant, but the support can evolve over time.
nnrec_w2	nnrec_w1 limited to 10 iterations.

Cross-Validation Algorithms

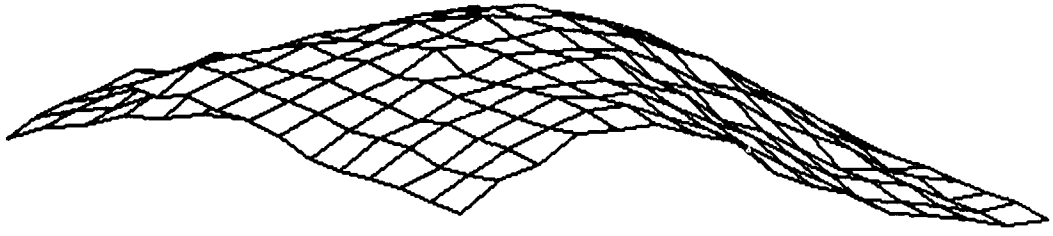
xvalid1	Calculates the regularization parameter for a 1-d image.
xvalid2	Calculates the regularization parameter for a block of 1-d images.
xvalid2d_1	Calculates the cross-validation function over a specified range of the regularization parameter for a given image.

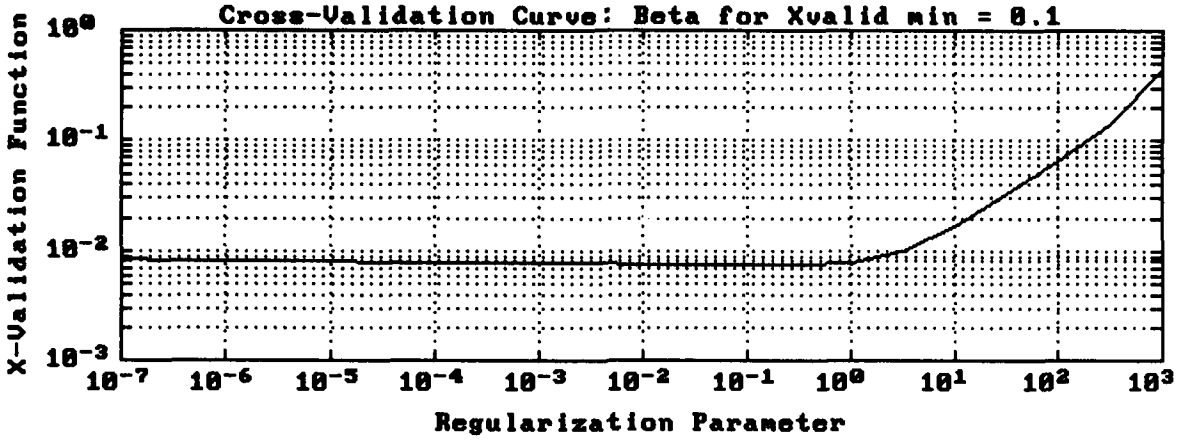
xv1d_l	Carries out the cross-validation Wahba sum using the svd of A.
xv1d_s	Carries out a truncated sum similar to xv1d_l, but generally shows increased variance in the cross-validation curve for small values of the regularization parameter, for 1-d images (caused by insufficient data points).

Object

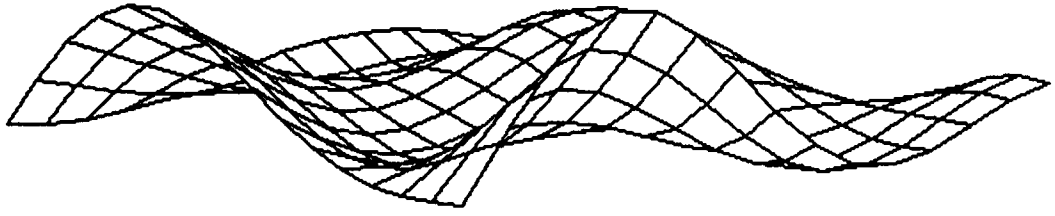


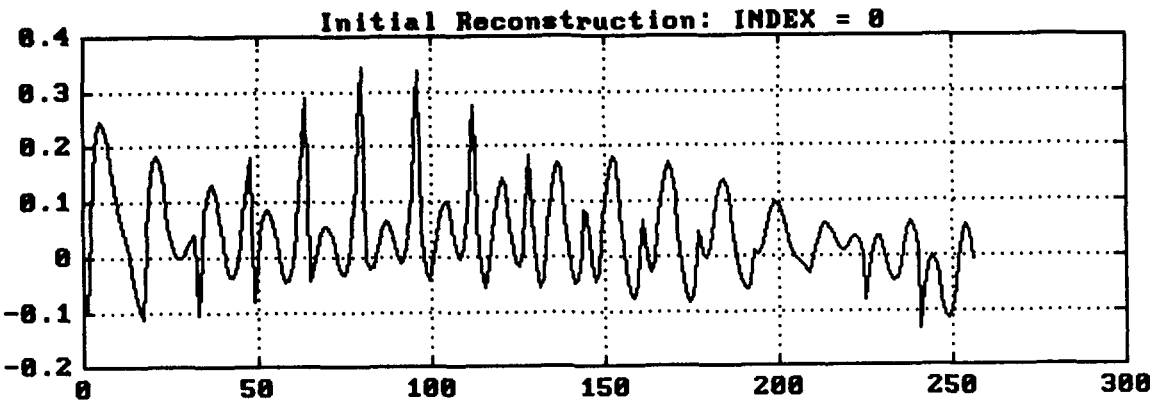
Image

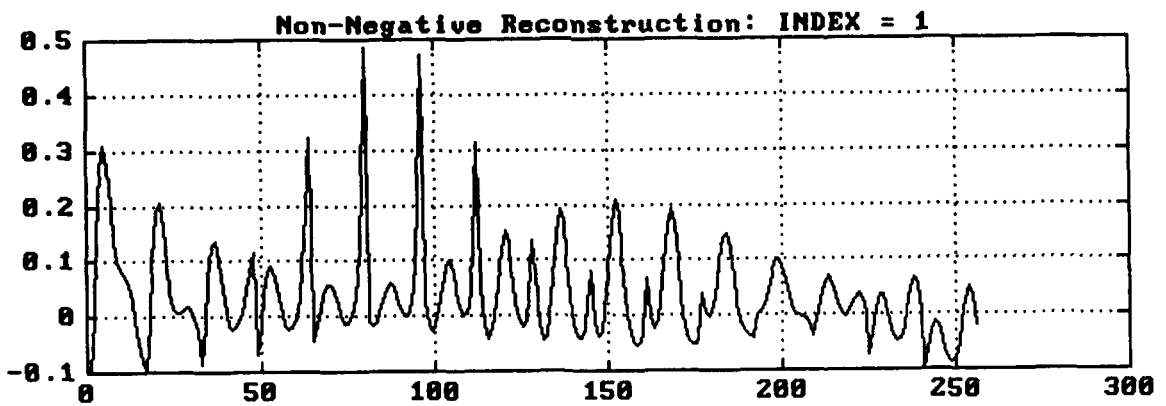


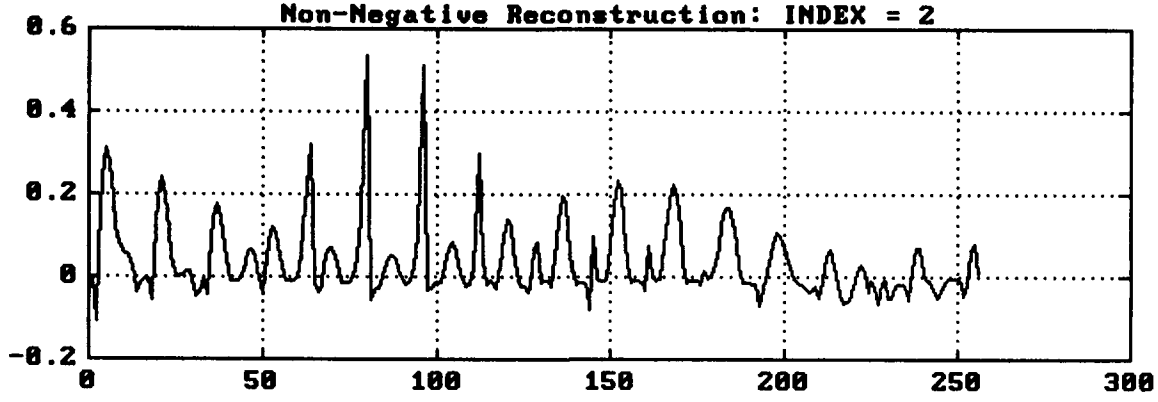


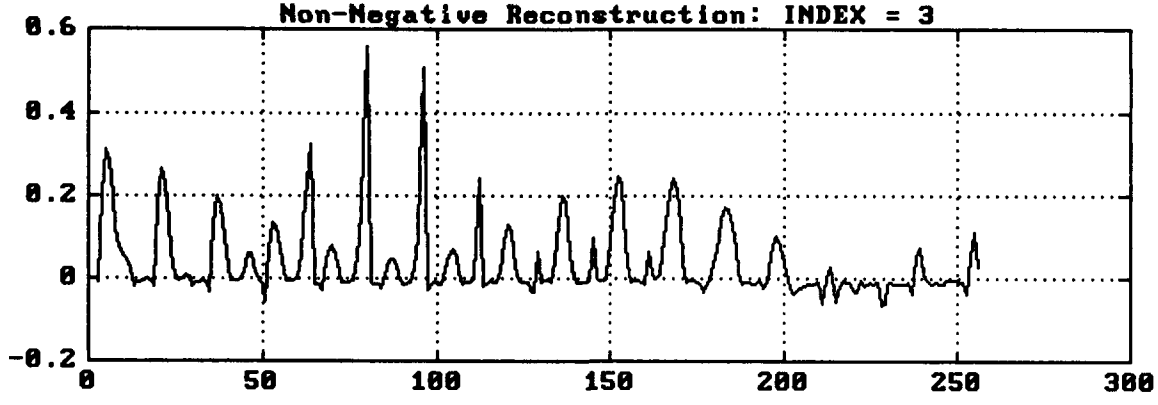
Reconstruction With Default Support: beta=0.1

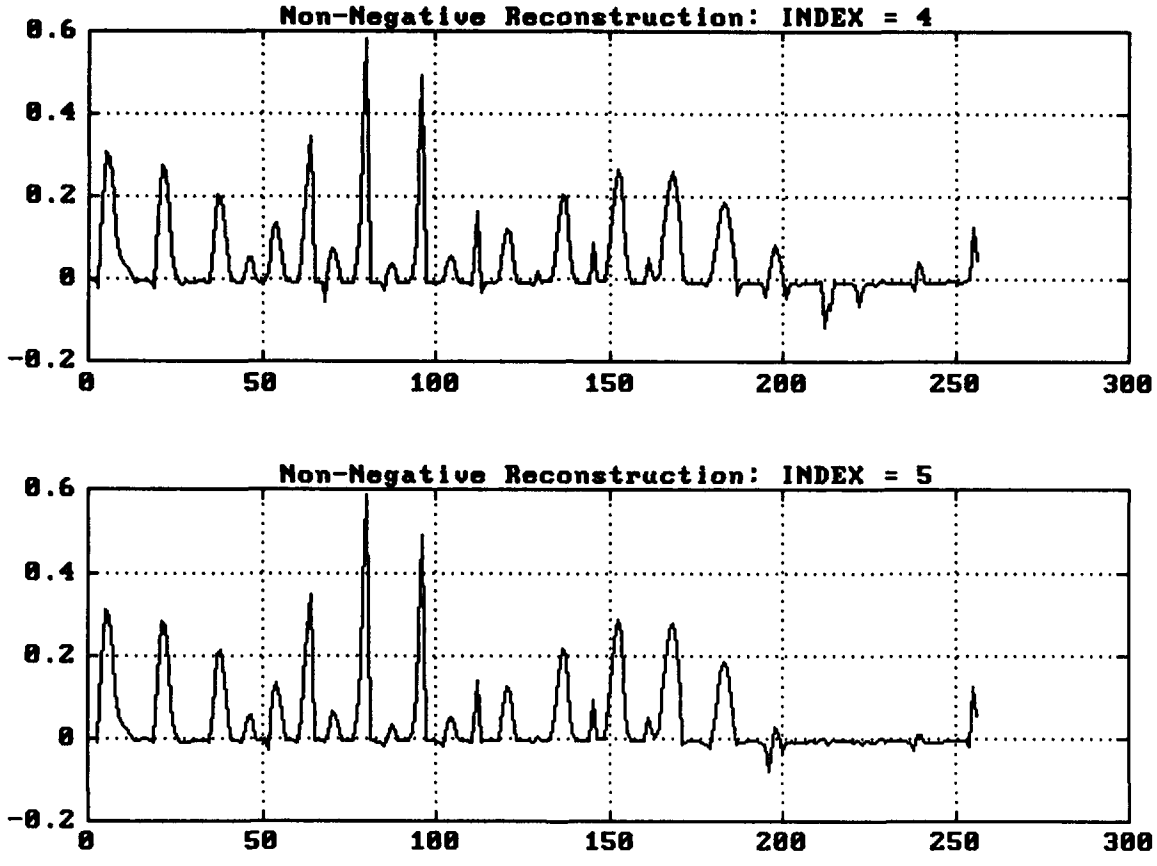


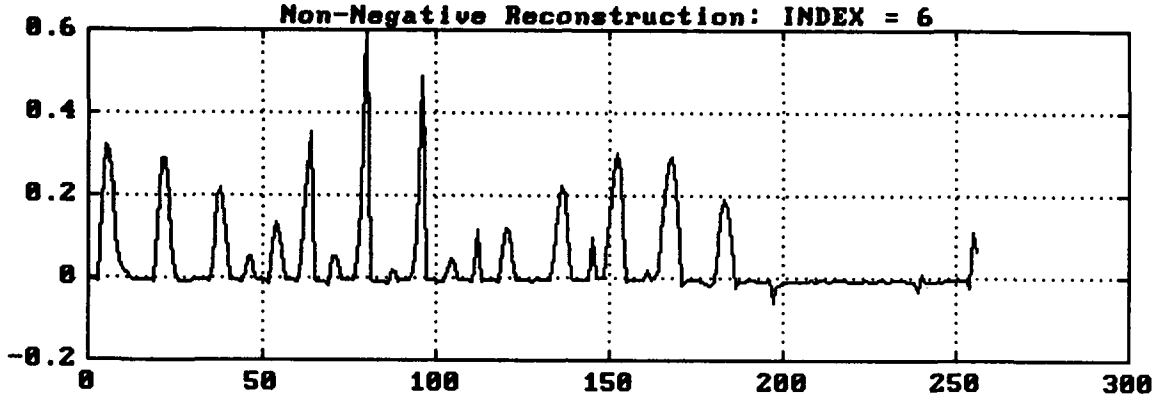

Initial Reconstruction: INDEX = 0
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
50
100
150
200
250
300

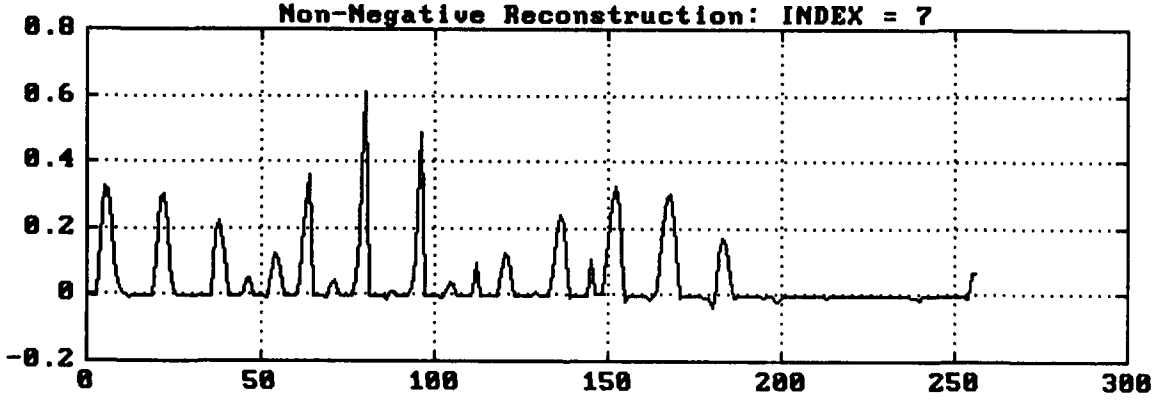

Non-Negative Reconstruction: INDEX = 1
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
50
100
150
200
250
300

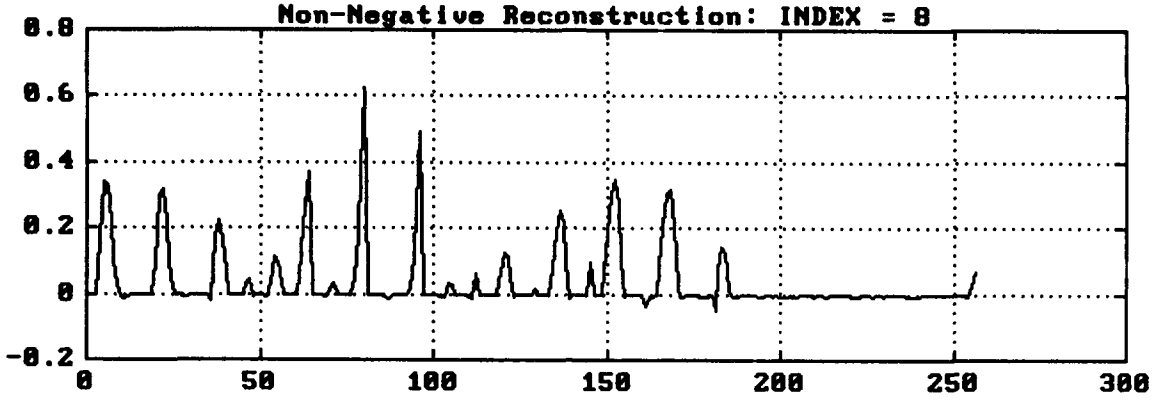

Non-Negative Reconstruction: INDEX = 2
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300

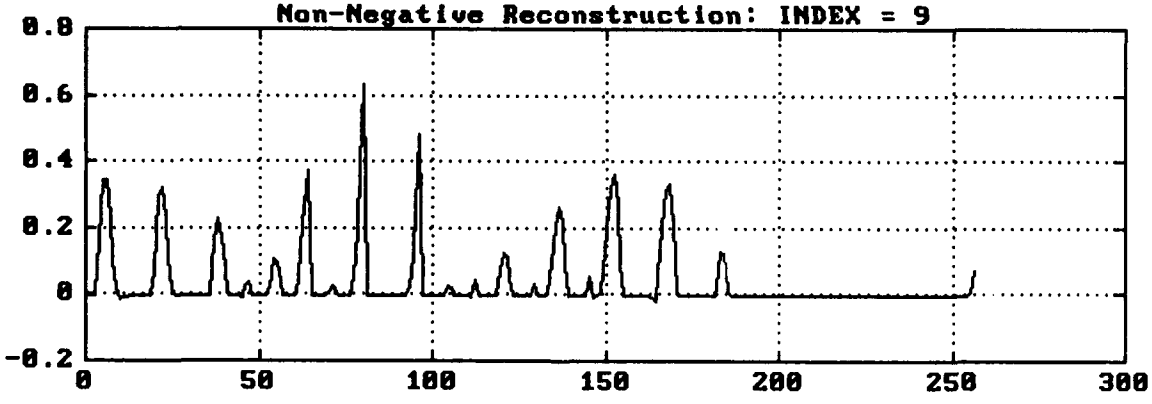

Non-Negative Reconstruction: INDEX = 3
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300

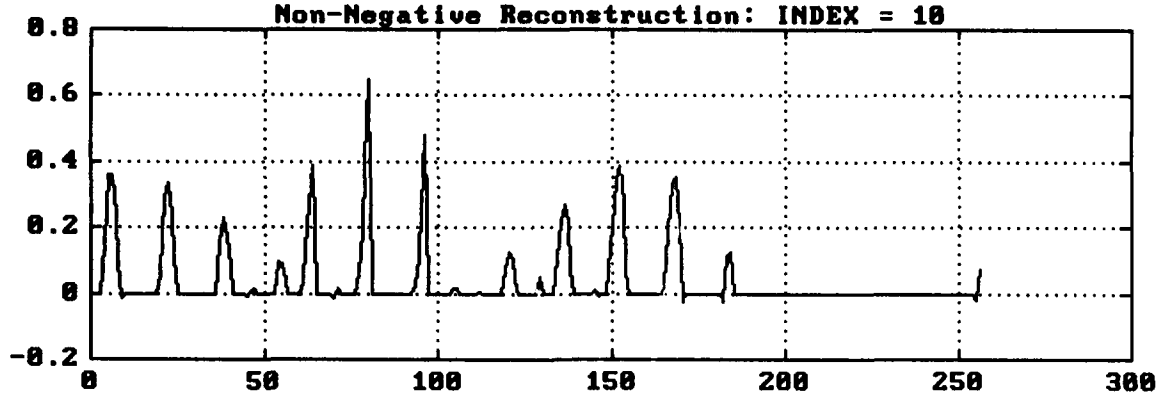

Non-Negative Reconstruction: INDEX = 4
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300
Non-Negative Reconstruction: INDEX = 5
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300


Non-Negative Reconstruction: INDEX = 6
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300

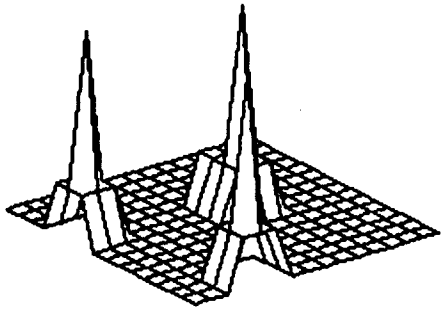

Non-Negative Reconstruction: INDEX = 7
0.8
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300


Non-Negative Reconstruction: INDEX = 8
0.8
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300

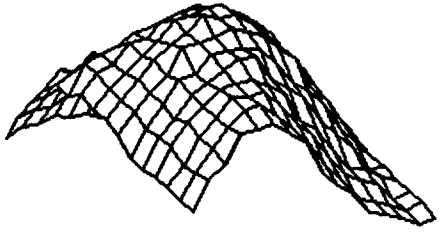

Non-Negative Reconstruction: INDEX = 9
0.8
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300


Non-Negative Reconstruction: INDEX = 10
0.8
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300

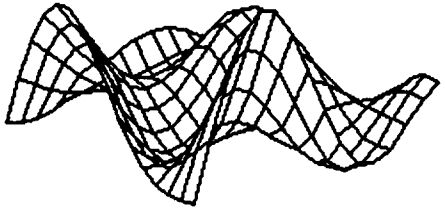
Object



Image

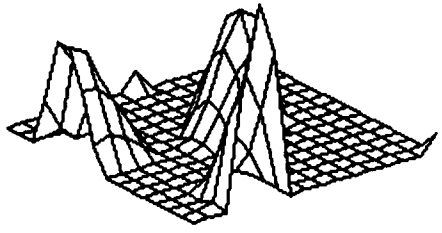


Default Support Reconstruction



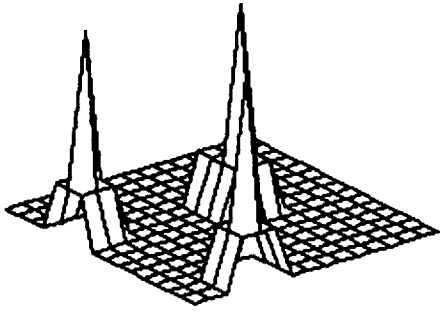
beta=0.1

Non-Negative Reconstruction

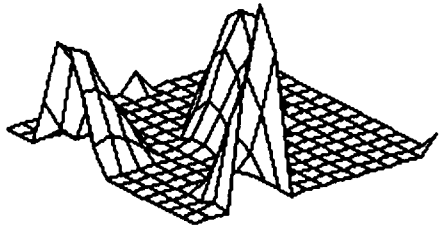


beta=0.03162

Object

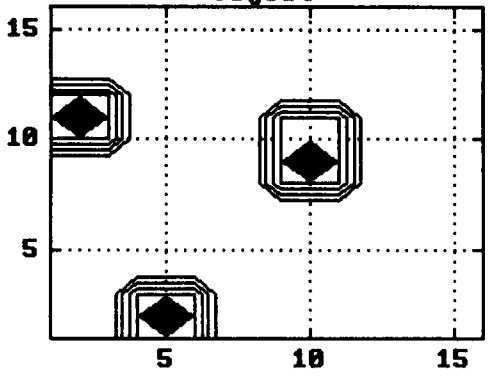


Non-Negative Reconstruction



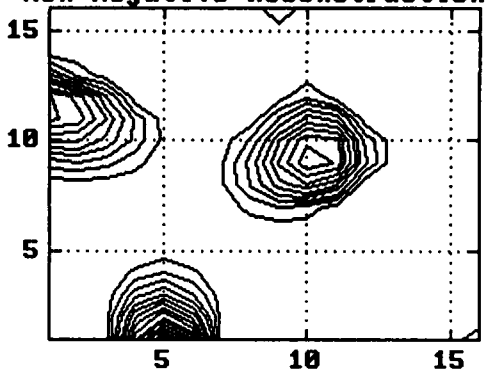
beta=0.1

Object



beta=0.03162

Non-Negative Reconstruction



APPENDIX C

CONTRACT PRESENTATIONS AND PUBLICATIONS

Presentations

J. B. Abbiss and H. J. Whitehouse, "Fast Regularized Reconstruction based on the Pseudo-Inverse of the Imaging Operator", Jet Propulsion Laboratory, Pasadena, California, 29 October 1990.

J. B. Abbiss and B. J. Brames, "Reconstruction of Sub-Pixel Detail from Defocused Images", SDIO/IST Workshop on Image Synthesis, Washington, D.C., 29-30 November 1990.

Publications

J. B. Abbiss, B. J. Brames and M. A. Fiddy, "Computational Aspects of Regularized Image Reconstruction", *Digital Image Synthesis and Inverse Optics*, Arthur F. Gmitro, Paul S. Idell and Ivan J. LaHaie, Editors, Proc. SPIE 1351, 85-96, 1990.

J. B. Abbiss, B. J. Brames and M. A. Fiddy, "Superresolution Algorithms for a Modified Hopfield Neural Network", IEEE Transactions on Signal Processing, Vol. 39, No. 7, 1516-1523, July 1991.

J. B. Abbiss and B. J. Brames, "Restoration of Sub-Pixel Detail using the Regularized Pseudo-Inverse of the Imaging Operator", *Advanced Signal Processing Algorithms, Architectures and Implementations II*, Franklin T. Luk, Editor, Proc. SPIE 1566, 365-375, 1991.

J. B. Abbiss, J. C. Allen, R. P. Bocker and H. J. Whitehouse, "Fast Image Reconstruction based on the Regularized Pseudo-Inverse of the Imaging Operator", *Inverse Problems in Scattering and Imaging*, Michael A. Fiddy, Editor, Proc. SPIE 1767, 93-111, 1992.

J. B. Abbiss, J. C. Allen, R. P. Bocker and H. J. Whitehouse, "Fast Regularized Deconvolution in Optics and Radar", Third IMA Conference on Mathematics in Signal Processing, University of Warwick, Warwick, UK, 15-17 December 1992.

Copies of the publications are appended.

COMPUTATIONAL ASPECTS OF REGULARIZED IMAGE RECONSTRUCTION

J B Abbiss, B J Brames, and M A Fiddy†

Spectron Development Laboratories, Inc., 3535 Hyland Avenue, Costa Mesa, California 92626
†Department of Electrical Engineering, University of Lowell, Lowell, Massachusetts 01854

ABSTRACT

Image restoration procedures are commonly unstable in the presence of noise, and some technique for restoring stability becomes essential. The methods of regularization theory are particularly appropriate for this purpose. A specific type of regularized solution is introduced in the general context of image reconstruction. A super-resolution problem is then considered from the point of view of the computational tasks involved, with particular reference to the estimation of certain key parameters and to implementations which increase the efficiency of the calculations. Parameter estimation is performed by weighted cross-validation. The improvement in efficiency is achieved through the exploitation of symmetries or cyclic properties inherent in the reconstruction operator. The concept of displacement rank is introduced and estimates made of the computational burden associated with various classes of regularized reconstruction matrices.

INTRODUCTION

We shall be concerned with the construction and computation of solutions to the problem of image reconstruction from noisy, incomplete or otherwise degraded data, when the imaging operator is known. The strongly smoothing operators which are typical of imaging or measurement systems are associated in the reconstruction process with an extreme sensitivity to noise or distortion in the data; small changes induced in the image can result in grossly differing reconstructions. The immediate cause of instability in the computational process is the ill-conditioning of the matrix representing the imaging operator.

The task of image reconstruction in fact belongs to the more general class of ill-posed inverse problems, which arise in many fields besides optics - for example in radar, geophysics, tomography or the extrapolation of band-limited signals. Suppose we are given the operator equation

$$Ax = y$$

where $A : X \rightarrow Y$. The problem of finding the solution x, given y, is said to be well-posed if for each y there exists a unique solution which depends continuously on the data; i.e., is stable under small changes in the data[1]. Otherwise, the problem is ill-posed. It should be emphasized that ill-posedness is inherent in the physical nature of the measurement process, not in the mathematical modeling of it.

Consider a typical image reconstruction problem. We suppose some object f , an element of a Hilbert space F , is imaged by a linear operator A into an element of another Hilbert space G , and that this image is perturbed by an unknown additive noise function which may arise in either the imaging or the observation process. The effect of the noise in practice will generally be to displace the image g out of the range of A - there will no longer be an element of F corresponding to g and the equation Af = g will have no solution. Clearly the problem is ill-posed.

This difficulty of a lack of continuous dependence can be overcome by one of the various techniques of "regularization". Essentially, the ill-posed problem is replaced by a related well-posed one, chosen to be physically meaningful and to possess the necessary properties of convergence and stability. Thus we may change the concept of a solution, or the Hilbert spaces or their topologies, or the operator itself. The technique we shall use belongs to the last category.

THE REGULARIZED RECONSTRUCTION

We impose physically reasonable constraints on the permitted solutions. If ε is a measure in the norm sense of the noise in the image (norm in the appropriate Hilbert space is denoted by $||\cdot||_H$) and if C is some constraint operator with E a known bound, we shall require that all possible reconstructions f' satisfy[2]

$$||g - Af'||_G \leq \varepsilon \quad \text{and} \quad ||Cf'||_F \leq E$$

C may be used, for example, to impose smoothness on the reconstruction or to weight the reconstruction support. If C is the identity operator, E is a bound on the norm of the reconstruction. We combine the constraints quadratically and minimize the functional

$$||g - Af'||_G^2 + \beta ||Cf'||_F^2$$

where $\beta = \varepsilon^2/E^2$. Note that smaller values of β are equivalent to demanding greater fidelity between the reconstruction and the data; greater values place more emphasis on the property of the reconstruction associated with C. In the present discussion we shall take C = I . The minimizer f_β can then be expressed in either of the forms

$$f_\beta = (A^*A + \beta I)^{-1} A^* g$$

or

$$f_\beta = A^* (AA^* + \beta I)^{-1} g$$

where A^* is the operator adjoint to A. The inverse of the bracketed operators will always exist, since the eigenvalues of the symmetric operators A^*A and AA^* are non-negative.

Weighting can be introduced into both of the spaces F and G through the definition of the inner product, and can be used to predispose the shape of the reconstruction to some prior expectation, or to reflect statistical information about the noise. A modified form of equation (1), incorporating the weighting matrices, now governs the reconstruction, which with the aid of Cholesky factorization can again be analyzed by SVD[4].

PARAMETER ESTIMATION USING WEIGHTED CROSS-VALIDATION:

To perform the computations in equation (1) or (3), we need estimates of the regularization parameter β and of the region in object space (the object support) into which the reconstruction will be made. In arriving at these estimates, the fullest use should be made of available a priori information about the object and the data-gathering process. Weighted cross-validation[5] provides a method for estimating a near-optimum value of the regularization parameter from the data alone. We demonstrate its use for this purpose and also for the estimation of the reconstruction support in the super-resolution problem.

ESTIMATING THE REGULARIZATION PARAMETER

It is possible to obtain a good estimate of the optimum regularization parameter from the data alone through the use of weighted cross-validation. The great merit of this technique is that it requires no knowledge of the object, and places very loose constraints on the image and noise, namely that the noise should be white and that the image be "very smooth" - this notion is defined precisely in Reference 5.

The idea behind weighted cross-validation is as follows. Suppose $f_{\beta,k}$ is the minimizer of

$$\sum_{\substack{j=1 \\ j\neq k}}^{N} \{(Ah)_j - g_j\}^2 + \beta ||h||_F^2$$

where the k^{th} data point has been omitted. We can now predict a value, $(Af_{\beta,k})_k$, for this k^{th} point and, for a given value of β, form the weighted prediction error $V(\beta)$:

$$V(\beta) = \sum_{k=1}^{N} \{(A f_{\alpha,k})_k - g_k\}^2 w_k(\beta)$$

where the $w_k(\beta)$ are weighting functions which represent the relative significance of the g_k in the prediction. They take the form

$$w_k(\beta) = \left[\frac{(AA^H + \beta I)_k^{-1}}{\text{trace}\{(AA^H + \beta I)^{-1}\}} \right]^2$$

It can be shown (Wahba) that the β which minimizes the expression above for $V(\beta)$ is an estimate of the minimizer of the true prediction error:

MATRIX OPERATORS AND THE SVD

The operator, the image and the reconstruction will consist in practice of finite arrays. We shall assume that the imaging operator has been recast (if necessary) as a two-dimensional matrix and that f and g are vectors. The finite-dimensional approximations to the Hilbert spaces F and G are K^N and K^M, N- and M-dimensional spaces of real or complex numbers. For matrices with complex entries, A^* becomes A^H and the solution is

$$\left.\begin{aligned} f_\beta &= (A^H A + \beta I_N)^{-1} A^H g \\ \text{or} \qquad f_\beta &= A^H (A A^H + \beta I_M)^{-1} g \end{aligned}\right\} \qquad (1)$$

To facilitate analysis and gain insight into the nature of the ill-posed problem we make use of the method of singular value decomposition (SVD)[3]. (There is also a significant gain in accuracy in using the SVD rather than in forming $A^H A$ directly.) We write for the SVD of A

$$A = U \Sigma V^H \qquad (2)$$

where

$$U^H U = V^H V = V V^H = I_N$$

and

$$\Sigma = \text{diag}\,(\sigma_1 \ldots \sigma_N), \quad \sigma_i \geq 0.$$

U consists of the N orthormalized eigenvectors u_i associated with the N largest eigenvalues of AA^H, and V consists of the N orthonormalized eigenvectors v_i of $A^H A$. The σ_i, the non-negative square roots of the eigenvalues of $A^H A$, are the singular values of A.

Then, from equation (1),

$$f_\beta = V \,\text{diag}\,(\ldots, \sigma_i/(\sigma_i^2 + \beta), \ldots)\, U^H g \qquad (3)$$

If $\beta = 0$, we have

$$f_\beta = A^+ g$$

where A^+ denotes the Moore-Penrose generalized inverse (familiarly called the pseudo-inverse) of A. Thus the operator in equation (1) can accurately be described as a regularized pseudo-inverse of A.

The regularizing role of β can clearly be seen in equation (3); for zero β, the elements of the diagonal matrix are simply the inverses of the non-zero singular values of A and can grow without bound as these tend toward zero.

$$T(\beta) = \sum_{j=1}^{N} \{(Af_\alpha)_j - \bar{g}_j\}^2$$

where $\bar{g}$ represents the ideal noiseless image

The estimator $V(\beta)$ can be expressed in the form

$$V(\beta) = \frac{\sum_{k=1}^{N} \{(AA^H + \beta I)^{-1} g\}_k^2}{[\,\mathrm{trace}\{(AA^H + \beta I)^{-1}\}]^2}$$

Thus to calculate $V(\beta)$ it is necessary to invert $AA^H + \beta I$. In the general case, $V(\beta)$ can be rewritten in terms of the SVD of A:

$$A = U\Sigma V^H, \quad \Sigma = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_N).$$

We find

$$V(\beta) = \sum_{k=1}^{N} \left(\frac{\hat{g}_k}{\sigma_k^2 + \beta}\right)^2 \Big/ \left\{ \sum_{k=1}^{N} \left(\frac{1}{\sigma_k^2 + \beta} \right) \right\}^2$$

where

$$\hat{g} = U^H G$$

Having carried out the SVD of A, this form has the advantage that recalculating $V(\beta)$ for a new value of β involves little more than 4N multiplications.

A further simplification can be made when A is a circulant matrix. In this case Af is a circular convolution, and the calculation can be carried out using a discrete Fourier transform (DFT), or, more efficiently, a fast Fourier transform (FFT). U in this case is the DFT matrix and $\hat{g}$ is the DFT of the data.

Although the problem of finding the regularization parameter would now be highly tractable, this approach is not usually applicable to superresolution for two reasons. The first reason is that in general the point-spread function in an imaging problem is neither a periodic function nor of finite extent, so that A cannot be a circulant. However, A can often be approximated as a circulant. The second consideration is relevant in spectral extrapolation: a circulant approximation to A enforces the same support bounds on the reconstruction as on the data g , so that only residual extrapolation is obtained outside the system bandpass. It might be possible in some cases to alleviate this problem by estimating the regularization parameter with a circulant approximation to A and then carrying out the actual reconstruction with the correct A.

ESTIMATING THE OBJECT SUPPORT

A number of approaches are possible to the problem of estimating the object support, which enters the calculation through the matrix A. Given the point-spread function of an imaging system, and possibly prior knowledge of the types of objects involved, one can sometimes form a good estimate of the outer boundary of the object support simply from the width of the image. An alternative strategy can be based on the fact that, for values of β around the optimum, the error estimator $V(\beta)$ is found to increase rapidly as the object support bound is reduced below the correct value. (The analytical behavior of $V(\beta)$ under these circumstances is still under investigation.) This suggests that we should choose the support bound associated with the minimum of the surface $V(\beta$, support bound).

An image with 5% Gaussian additive noise was formed of the object shown dashed in Figure 1. The object support was 17 units wide. A sequence of cross-validation curves was calculated over the range $10^{-9} \leq \beta \leq 10^{8}$ for reconstruction supports of 1 to 25 pixels. The resulting surface is shown in Figure 2. The minimum occurs at the point corresponding to $\beta = 0.1$, support = 17. A set of values of $V(\beta)$ was then calculated for $\beta = 0.1$ and for a series of reconstruction supports consisting of the correct support bound of 19, but differing by the deletion of successive internal single points. Thus the reconstruction was forced to zero on one of the 17 pixels in each case. Figure 3 shows the resulting values of $V(\beta)$ plotted against the corresponding deleted pixels. The positional dependence approximately conforms to the correct object support.

DISPLACEMENT RANK AND FAST MATRIX-VECTOR MULTIPLICATION

Inversion of an N×N matrix takes in general $O(N^3)$ operations. If the matrix is Toeplitz the number of operations needed falls to $O(N^2)$. If the imaging operator is shift-invariant and the numbers of samples and sampling rates (for unit magnification) are the same in image and reconstruction spaces, then the imaging matrix A is Toeplitz. However, it is often the case that A is not Toeplitz, but nevertheless retains much Toeplitz-like structure. This is true, for example, if we assume that the point-spread function is shift-invariant, but that A is rectangular, or that the sampling rates in image and reconstruction spaces are unequal, or that columns are deleted from A to restrict the reconstruction support in some appropriate way. In all of these cases, A retains some regularity in structure which one would expect could be used to make computations involving A more efficient.

The concept of displacement rank was introduced to quantify these ideas[6]. The SVD then provides a mechanism for constructing approximations to A in a form in which its special properties can be exploited[7]. We define first the N×N downshift matrix Z:

$$Zx = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} 0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix}$$

The displaced difference of an arbitrary N×N matrix B is defined as $B\text{-}ZBZ^H$. By means of the SVD, the displaced difference can be written

$$B - ZBZ^H = \sum_{i=1}^{N} x_i y_i^H \quad .$$

If the SVD is truncated at r terms, we obtain a reduced displacement rank r approximation to B. An equation of the above form has the unique solution[6]

$$B = \sum_{i=1}^{r} L(x_i)U(y_i) \quad ,$$

where L is a lower triangular Toeplitz matrix whose first column is x_i and U is an upper triangular Toeplitz matrix whose first row is y_i^H. Then multiplication of a vector by B can be accomplished in 2r convolutions. For sufficiently small r, a significant saving of effort will results from performing these with the FFT, the number of operations falling from $O(N^2)$, for the direct computation of Bg, to $O(2rN\log_2 N)$. In evaluating the product, $U(y_i)g$ yields the right half of the convolution between the top row of $U(y_i)$ and g. Similarly, $L(x_i)g$ is the left half of the convolution between the bottom row of $L(x_i)$ and g. Note that, for a Toeplitz matrix, the displaced difference contains non-zero elements only in the leading row and column, is of rank 2 and can be represented by the sum of two products of lower and upper triangular Toeplitz matrices.

The triangular matrices in the expansion for B can be written

$$L(x_i) = [x_i \mid Zx_i \mid \ldots \mid Z^{N-1}x_i]$$

and

$$U(y_i) = [y_i \mid Zy_i \mid \ldots \mid Z^{N-1}y_i]^H$$

The displaced difference defined above is most relevant when large blocks of B are displaced in the (1, 1) direction. In many cases of interest (arising in practice from more general sampling schemes) we encounter matrices having blocks displaced in other directions, and the lowest rank representation will be obtained when the displacement reflects the structure of the matrix. Let us define a generalized displaced difference as

$$D_{pq} = B - XBY^H$$

where $X = Z^p$ and $Y = Z^q$.

B can now be expanded into a sum of products of lower and upper triangular matrices (not now in general Toeplitz):

$$L(x_i) = [x_i \mid Xx_i \mid \ldots \mid X^{N-1}x_i]$$

and

$$U(y_i) = [y_i \mid Yy_i \mid \ldots \mid Y^{N-1}y_i]^H$$

DISCUSSION

The number of operations necessary to form the product Bg is approximately $2N\{r+(r+1)\log_2 N\}$ where we have assumed g is real and that the FFT involves $O(N\log_2 N)$ operations. Note that for each term LUg, two FFTs must be performed since Ug yields the right half of a convolution and L(Ug) yields the left half. It is necessary to inverse Fourier transform after performing the first convolution, select the appropriate subset of the data and then re-transform to perform the convolution with L. For $r = 2$ there is a computational gain over the direct matrix-vector multiplication for $N \geq 35$ and for $r = 4$ there is a gain for $N \geq 70$. For two-dimensional non-separable point-spread functions, and $r = 4$, this threshold would be reached by a 9×9 image, since the imaging operator would then be written out as an 81×81 matrix. The computing cost for a 32×32 image in this case drops to about one-tenth that of the direct multiplication (order 32^4 floating-point operations).

Modifications of the Toeplitz imaging matrix A are induced by changing the sampling scheme in image or reconstruction space; for example, by periodic deletion of rows or columns to modify the data or reconstruction supports. Note that we are particularly concerned here with computing the product Qg, where Q is the regularized pseudoinverse of A. We recall that $Q = (A^HA+\beta I)^{-1}A^H$. An upper bound on the displacement rank of Q is given by[7]

$$r(Q) \leq 2r(A)+3,$$

which places a bound on the number of operations required to form Qg. We note also that for isoplanatic imaging systems and equal sampling intervals in data and reconstruction space, A is Toeplitz and has displacement rank 2; $(A^HA+\beta I)$ can then be inverted in $O(4N^2)$ operations.

It may not be convenient from a hardware point-of-view to take the data with the same spatial sampling that one desires to carry out the reconstruction. If one forms A from a Toeplitz matrix T by removing alternate rows, then the sampling interval in the image is twice that in the reconstruction. For clarity, let us keep the same number of points in image and reconstruction space, so A is N×N. Then in general the displacement rank of A can be as large as N. However, we have not properly exploited the structure in A, which is along the direction (1,2), rather than (1,1), as in a Toeplitz matrix. If we consider instead $D_{12}(A) = A-ZA(Z^H)^2$, we find that $D_{12}(A)$ is of rank ≤ 3. In general, if we form A_1 by deleting every n^{th} row, $D_{1n}(A_1) = A_1 - ZA_1(Z^H)^n$ will have rank $(n+1)$. A_1 can again be expanded into a sum of products of lower and upper triangular matrices. Note that in this case the U matrices are no longer Toeplitz. However, because the product $U(y_i)g$ yields every n^{th} point of the right half of the convolution of y_i and g, we can still use the FFT to perform the convolution, discarding the undesired points. If, on the other hand, every n^{th} column had been deleted to form A_2, the appropriate displacement

difference would be $D_{n1}(A_2) = A_2 - Z^n A_2 Z^H$, and a similar procedure could be used in computing the product Lh, where h = Ug. (The data points in h have also to be interleaved with zeros in appropriate places.)

Unfortunately, for A formed of alternate rows of a Toeplitz matrix the rank of $D_{12}(Q)$ and $D_{21}(Q)$ are less than or equal to the full rank of Q even though $D_{12}(A) = 3$. It seems that one cannot generally expect the reduced displacement rank representation to offer greater computational efficiency unless the sampling rate is the same in image and reconstruction spaces. However, in some particular cases of dissimilar sampling rates, low displacement ranks for Q have been obtained. For example, in experiments with imaging matrices of $(\sin x/x)^2$ type and constructed with the sampling rate in reconstruction space twice that in image space, it was found that $D_{21}(Q)$ was of much less than full rank. The features of the matrix which are responsible for this property are being investigated.

CONCLUSIONS

A regularized solution to the image reconstruction problem has been discussed and the use of weighted cross-validation demonstrated in the estimation of the regularization parameter and the reconstruction support. We have also shown how low displacement rank representations can be used to improve computational efficiency for imaging matrices of practical interest. These techniques are not limited to the example from optical physics considered here, but should be widely applicable to problems in the general area of information processing.

ACKNOWLEDGEMENTS

The authors would like to acknowledge stimulating and informative discussions with Drs. Jeff Speiser, Harper Whitehouse and Keith Bromley of the Naval Ocean Systems Center, San Diego. The work was supported by SDIO/IST and managed by ONR.

REFERENCES

1. Tikhonov, A. N. and Arsenin, V. Y., *Solutions of Ill-Posed Problems*, V. H. Winston and Sons, Washington D.C. (1977).

2. Miller, K., "Least squares methods for ill-posed problems with a prescribed bound", SIAM J. Math. Anal., 1, 1, 52-74 (1970).

3. Golub, G. H. and Reinsch, C., "Singular value decomposition and least squares solutions", Numer. Math., 14, 403-20 (1970).

4. Abbiss, J. B. and Earwicker, P. G., "Compact operator equations, regularization and super-resolution", Proceedings, IMA Conference on 'Mathematics in Signal Processing', University of Bath, Bath, U. K., 17-19 Sept 1985.

5. Wahba, G., "Practical approximate solutions to linear operator equations when the data are noisy", SIAM J. Numer. Anal. 14, 4, 651-667 (1977).

6. Kailath, T., Kung, S. Y. and Morf, M., "Displacement rank of matrices and linear equations", J. Math. Anal. and Appns, 68, 395-407 (1979)

7. Speiser, J. M., Whitehouse, H. J. and Allen, J. C., "Fast matrix-vector multiplication using displacement rank approximation via an SVD", Proceedings, Second International Workshop on SVD and Signal Processing, University of Rhode Island, Rhode Island, 25-27 June 1990.

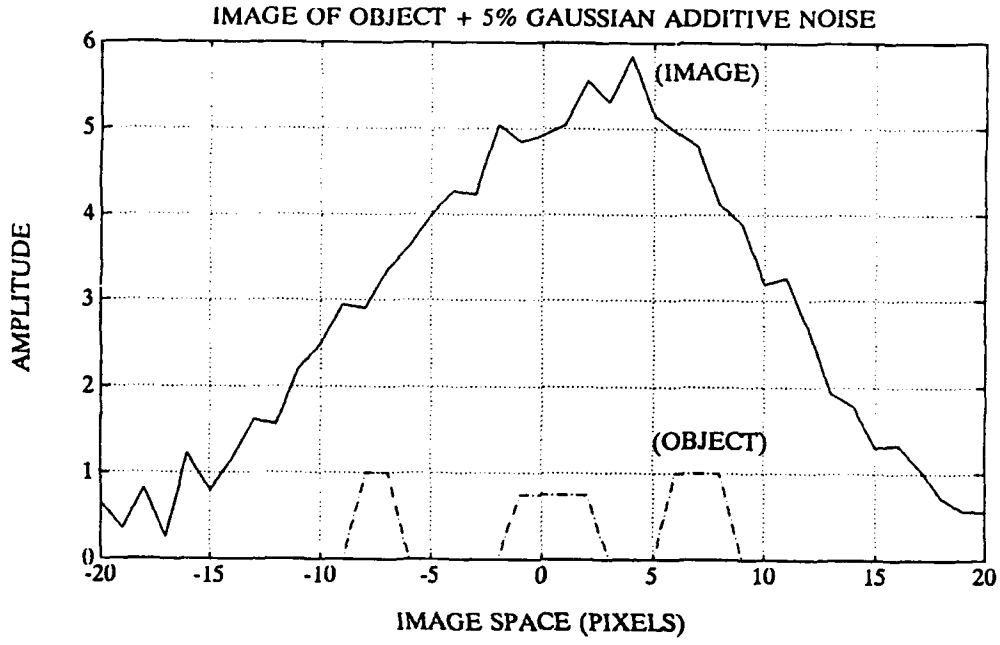


Figure 1: A scaled image of an object 17 pixels wide. The image contains 5% Gaussian additive noise.

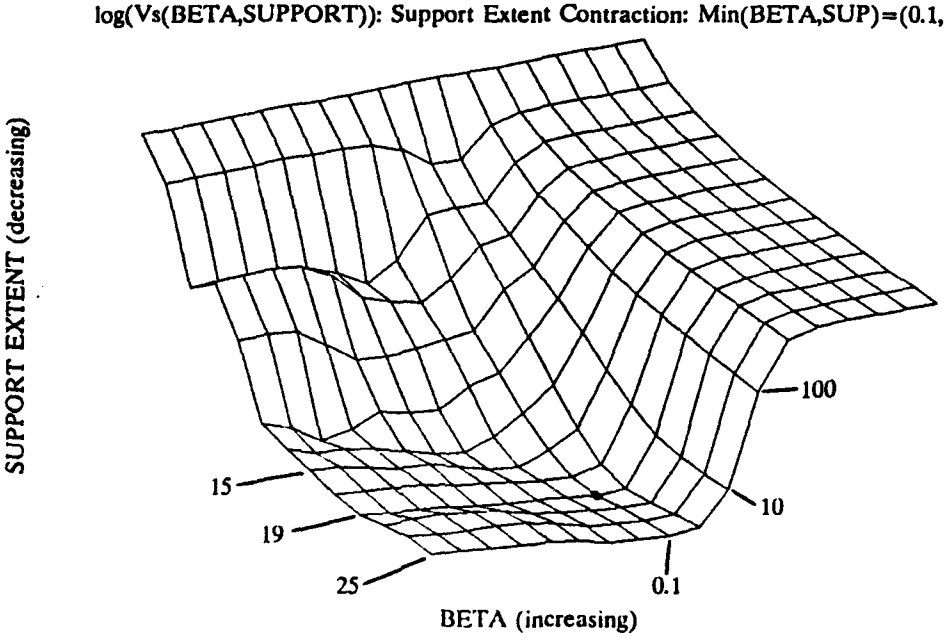


Figure 2: Cross-validation prediction error V(β, support). The regularization parameter ranges from $1e^{-9}$ to $1e^{8}$, and reconstruction support widths range from 25 to 1 pixel. The minimum occurs for β = 0.1 and support width = 19 pixels. The object support was 17 pixels wide.

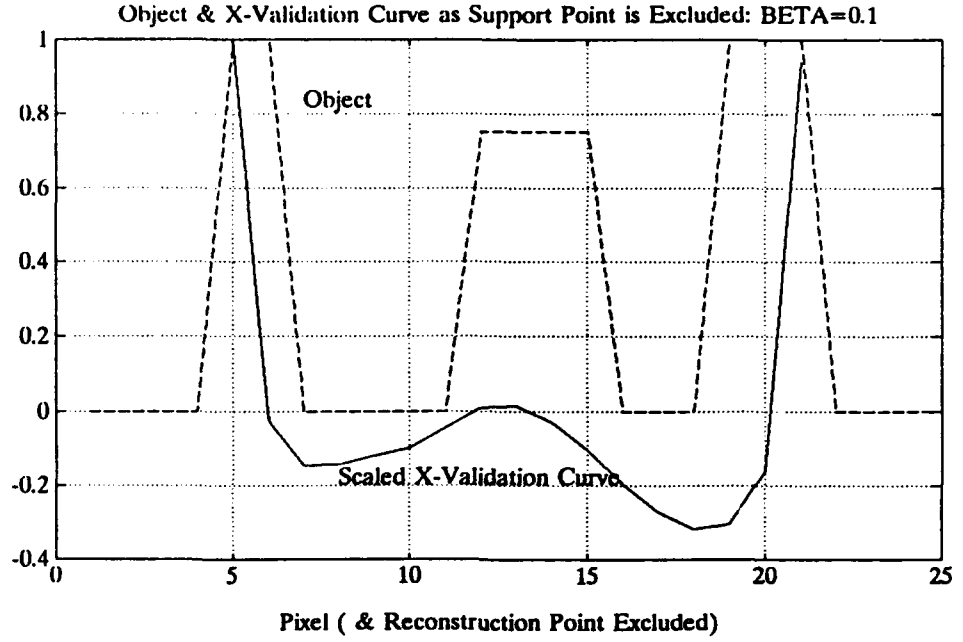


Figure 3: Object and superimposed cross-validation prediction error $V(\beta = 0.1, \text{excluded\_point})$. Observe that V reflects some of the internal support structure of the object and gives a good estimate of the object width.

Superresolution Algorithms for a Modified Hopfield Neural Network

John B. Abbiss, Bryan J. Brames, and M. A. Fiddy, *Member, IEEE*

Abstract—**The purpose of this paper is to describe the implementation of a superresolution (or spectral extrapolation) procedure on a neural network, based on the Hopfield model. This was first proposed by Abbiss *et al.* [1]. We show the computational advantages and disadvantages of such an approach for different coding schemes and for networks consisting of very simple two state elements as well as those made up of more complex nodes capable of representing a continuum. With the appropriate hardware, we show that there is a computational advantage in using the Hopfield architecture over some alternative methods for computing the same solution. We also discuss the relationship between a particular mode of operation of the neural network and the regularized Gerchberg–Papoulis algorithm.**

I. Introduction

THERE are several models of neural networks, each of which has a structure based loosely on our view of biological nervous system components [2]. A neural network architecture is one consisting of a very large number of simple processing elements densely interconnected by a set of weighted links. Each processing element updates its state by comparing the sum of its inputs with a prescribed threshold. The study of the properties of neural networks is a subject still somewhat in its infancy, and current hardware limitations reduce their practical impact. Indeed, it has been suggested by Anderson and Rosenfeld [3] that they may not become useful until inexpensive special purpose parallel hardware is available. Should that hardware be available, the question remains as to how one would make best use of a neural computer; i.e., how one should program or "train" it to perform the tasks required. The hope is that some problems for which it is difficult to find satisfactory algorithmic solutions might be amenable to this kind of computing architecture, which can organize itself and learn what it is expected to accomplish.

One anticipated use of neural networks is in autoassociative memory and in image (or signal) classification, recognition or understanding applications; these are applications that we believe the human brain is particularly good at while current algorithms implemented largely on serial machines still leave much to be desired. For most neural networks, their learning or restoration capabilities can be expressed in terms of the minimization of some appropriate energy or cost function. One of the objectives of this paper is to take an established algorithm in image reconstruction and identify those aspects of it that can be related to the programming requirements that would be necessary for implementation on a Hopfield neural computer. From the analysis of such a network applied to solve a problem for which the cost function is well defined, one might be able to assess their use for the solution of a wider class of optimization problems.

The Hopfield network is a fully connected network in the sense that any one of the processing elements is connected to every other one. This contrasts with layered networks, such as a multilayered perceptron (MLP), in which processing elements are arranged with connections only between neighboring layers. This difference in topology is accompanied by differences in the thresholding functions and in the procedures to find the connection strengths. The Hopfield network operates iteratively; the connection strengths are assigned and specify a cost function which the iterative procedure minimizes. The MLP is a one-pass network once the connection strengths have been "learned" by the minimization of an error function which quantifies the difference between the current and desired output states.

It was pointed out by Jau *et al.* [4] that some iterating image restoration processes are mathematically very similar to autoassociative memory; indeed if the input information is incomplete, it can be considered as a key pattern to an associative memory. Since the approach to image restoration presented here was first proposed [1], there have been other related studies which we mention here. Zhou *et al.* [5] considered an energy function identical to (6) in order to specify network interconnection strengths. Their application was the restoration of grey level images degraded by a shift invariant FIR blur function and additive noise. Grey level information was coded by a simple sum of binary elements and the network was serially updated with a stochastic thresholding rule to avoid getting trapped in local minima of the energy function. Jang *et al.* [6] utilized the optimization properties of the Hopfield network in order to estimate a matrix inverse. This is clearly important in image restoration problems, as indicated by (8). Full grey level representation is assumed and

Manuscript received April 25, 1989; revised August 23, 1990. This work was supported in part by the SDIO/IST and managed by ONR.

J. B. Abbiss and B. J. Brames are with Spectron Development Laboratories, Inc., Costa Mesa, CA 92626.

M. A. Fiddy is with the Department of Electrical Engineering, University of Lowell, Lowell, MA 01854.

IEEE Log Number 9144725.

1053-587X/91/0700-1516$01.00 © 1991 IEEE

a differential mode of implementation applied; they point out that this method is similar to a steepest descent method but with a nonlinear thresholding step at each iteration. Bai and Farhat [7] also considered a cost function similar to that in (6) to recover images from limited Fourier data. In their approach there was an additional constraint on the norm of the derivative of the estimate as well as on the norm of the estimate itself. The increment added to the current estimate was weighted by a "gain" factor prior to thresholding, which was chosen to ensure that the network energy function decreased at each step. Their reconstructions showed advantages for low-signal-to-noise ratio situations and are being implemented on optoelectronic hardware. Winters [8] considers a norm minimization as expressed by (6) but without any explicit regularization term included. The minimization of his cost function is achieved by the penalty method which requires that a large positive value is added to the cost function, wherever a nonlinear inequality constraint is not satisfied. An adaptive penalty function allows one to avoid local minima and this complete procedure can be mapped onto the Hopfield energy function. Results showed that the reconstructions were robust against noise and could be implemented in microseconds on an analog electronic network, as compared to several hours on a minicomputer.

Using a Hopfield network, our interest is in an application for which a solution state evolves through the minimization of some specific cost or energy function. Once the energy function is defined, one can determine the appropriate connection strengths in order that the energy function associated with the network is the same as that of the problem under consideration. A key feature of a net of this type is its construction from a set of simple processors, each of whose states is determined by a thresholding operation applied to a sum of weighted inputs from other processors or nodes. The properties of the network as a whole are determined by the thresholding function used, and by the pattern and strengths of the connections between the processing elements.

II. The Hopfield Network with Two State Elements: Theoretical Framework

The Hopfield neural model [9], [10] allows one to specify a set of desired memories as minima of a configurational energy of the network. We assume that the network consists of N processing elements each of which has two states and each of which has a thresholding operator that determines the states of the element from the total input to that element.

Given an initial starting configuration or state of the network, each processor or "neuron" updates its state according to a threshold rule of the form

$$\text{if } \sum_{j=1}^{N} T_{ij} v_j > 0 \text{ then } v_i = 1; \quad \text{otherwise } v_i = -1 \qquad (1)$$

where the elements of the connection matrix T are formed according to the rule

$$T_{ij} = \sum_{\mu=1}^{M} v_i^{\mu} v_j^{\mu} \qquad (i, j = 1, 2 \cdots N) \qquad (2)$$

and the v_i^{μ} are the elements of the M memory vectors, v^{μ}, to be stored. The v_i^{μ} can take the values ± 1, and it is assumed that the diagonal term, T_{ii}, is zero in the Hopfield model.

This thresholding rule can be applied in series (asynchronously) or in parallel (synchronously). In the serial mode, the rule is applied sequentially to the nodes of the network, the state of the network being updated after each operation. In the parallel mode, the current network state remains unchanged until the thresholding operation has been applied to every node. The configurational energy function for the network has the form [9], [10]

$$E = -\tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} v_i v_j = -\tfrac{1}{2} v^T T v \qquad (3)$$

where superscript T denotes transpose. Serial thresholding will always minimize this energy function, provided that the T_{ii} are nonnegative. If M is sufficiently small, this state will correspond to the memory closest in Hamming distance to the state in which the network was started. Parallel thresholding results in either convergence to a stable state or oscillation between two states [11].

This iterative scheme can be expressed more concisely and modified to permit biasing of the neuron inputs in the following manner. We let the state of the network after the nth iteration be described by the N element vector

$$v^{(n+1)} = U(Tv^{(n)} + b)$$

where U is the threshold operation, T denotes the matrix with elements T_{ij}, a superscript denotes iteration number, and b is the bias vector. The bias vector incorporates boundary conditions such as image data; it effectively shifts the decision threshold for each element. In this case the energy function minimized by the network is of the form [10]

$$E = -1/2\, v^T T v - b^T v. \qquad (4)$$

The change in energy for a change in the state of one neural element from v_k to $v_k + \Delta v_k$ is

$$\Delta E_k = -\Delta v_k [(Tv + b)_k + \tfrac{1}{2} T_{kk} \Delta v_k].$$

Taking T_{kk} to be zero ensures that the change in energy is always negative, since the term in the brackets above then has the same sign as Δv_k. If T_{kk} is nonzero, the term in the bracket will have the same sign as Δ_k provided T_{kk} is positive, and then E is guaranteed to reduce; the consequences of varying T_{kk} were explored in an earlier publication and verify the expected behavior [12]. If two (or more) neurons change state simultaneously, the change in E contains terms involving products of the form $-1/2\, T_{kl} \Delta v_k \Delta v_l$ (or these plus higher order terms if more neurons change), the sign of which can vary.

The two state representation is too limited for a one-to-one mapping between elements and signal or image samples, in most cases. However, these simple elements can be taken in groups to represent grey levels through a variety of coding schemes. Alternatively, either analog or more complex digital processing elements could be used to directly represent a grey level.

III. The Superresolution Problem

There are many applications that require the restoration of a signal or image from a limited discrete data set; for example, samples of the spectrum of a function, or of its low or band-pass filtered image. An important *a priori* assumption for work in super resolution is the fact that most objects to be imaged are of compact support. This leads to the well-known result that their spectra are band-limited functions. In principle, therefore, one might hope to extend limited spectral data by means of analytic continuation. This procedure is notoriously unstable in the presence of noise and does not provide a practical solution to the problem. One has infinite freedom in interpolating and extrapolating limited sampled data; hence, one is forced to approach super resolution from an optimization point of view [13]. The best that one can hope to achieve is the specification of a cost or energy function which possesses a unique minimum and is designed to incorporate whatever constraints and *a priori* knowledge might be available to help limit the set of possible solutions to the problem, while retaining desirable and necessary solution characteristics. Examples of constraints include data consistency, support consistency and, perhaps, positivity. The objective of the superresolution process is to obtain a final image that has a higher spectral or spatial frequency content than the original data set, as a direct consequence of incorporating the prior knowledge available into the cost function. It is a matter of taste, to a large extent, how one designs a cost function in order to obtain a desirable solution to the problem; i.e., a superresolved signal or image with acceptable properties. The superresolution problem is thus transformed into one of determining the (global) extremum of a cost function on the assumption that this solution is optimum.

One of the early successes of a neural network was to find a good approximation to the traveling salesman optimization problem [14]. The superresolution optimization problem can be mapped onto a neural network in two distinct ways. One is to train network using a data base of superresolved images [15]–[17], the other is to relate the cost function associated with a given network to the chosen superresolution cost function. It is the latter approach that we adopt here.

IV. Superresolution and Spectral Estimation

Most signal or image recovery problems can be described by linear equations of the form

$$g(x) = \int A(x, y) f(y)\, dy$$

where A is the system spread function or the Fourier transform kernel, for example. The interpretation of the data $g(x)$ to obtain information about the object $f(y)$ requires the solution of a linear inverse problem. This is equivalent to finding the solution of a Fredholm integral equation of the first kind. It is well known that small fluctuations in the data $g(x)$ can lead to very large fluctuations in the estimate of the unknown function $f(y)$. This is a manifestation of the ill-posed nature of the problem (the inverse of the operator A is not generally continuous) and some degree of regularization is required in order to determine stable and meaningful solutions. One usually proceeds by assuming that the desired solution belongs to the space F of (possibly weighted) L^2 functions, the regularization restricting that solution to conform to any *a priori* knowledge available about the object whose enhanced image is sought.

In practice, the solution is determined from a finite set of samples of $g(x)$, and the data vector g is expressed by

$$g = Af + n$$

where A is the imaging operator and n represents an additive noise component; A explicitly contains the support constraint of f, which is assumed to be known or estimated *a priori*. These limited data can be regarded as a noisy finite set of bounded linear functionals of f.

A data-consistent solution exists, however, which is a solution of minimum norm. This solution is the data-consistent $\hat{f}$ which minimizes $\|\hat{f}\|^2$, where $\|\ \|$ denotes norm.

The solution to this minimization problem can be written

$$\hat{f} = \sum_{i=1}^{N} (g, v_i) u_i / \alpha_i \tag{5}$$

where N is the number of image data points and the α_i, u_i and v_i are the singular values, singular functions, and singular vectors, respectively, pertaining to the operator A: $Au_i = \alpha_i v_i$; $A^* v_i = \alpha_i u_i$. The singular values tend to zero as i increases, leading to an instability in the estimator. If the first $N_i \leq N$ singular values are dominant, then the remainder may be neglected, but only at the expense of loss of resolution in $\hat{f}$.

Thus, this solution is ill-conditioned but stability can be restored by relaxing data consistency; hence, we minimize the cost function

$$E = \|A\hat{f}' - g\|^2 + \beta \|\hat{f}'\|^2. \tag{6}$$

The estimate is

$$\hat{f} = \sum_{i=1}^{N} (g, v_i) u_i \alpha_i / (\alpha_i^2 + \beta) \tag{7}$$

where the regularization parameter β is chosen to achieve a compromise between resolution and stability, and usually requires some adjustment in order to establish its optimal value. As β tends to zero, the solution becomes more

data consistent. The minimizer of this cost function can be computed directly in matrix form, namely,

$$\hat{f} = [A^*A + \beta I]^{-1}A^*g \tag{8}$$

where, for a real-valued matrix, A^* becomes A^T, and I represents the identity matrix.

An alternative approach to estimating the object is to consider the minimization of the cost function $\|f - \hat{f}'\|^2$ using a trigonometric polynomial of the form [13]

$$\hat{f} = \sum_{k=1}^{N} d_k \phi_k$$

where the ϕ_k form a basis in the data space F and the optimal d_k satisfy

$$\sum_{m=1}^{N} [(\phi_m, \phi_n)_F + \beta\delta_{nm}]d_m = G_n \tag{9}$$

and the G_n are Fourier data corresponding to the low-pass filtered image g.

It is worth pointing out that in the space F that incorporates the known support constraint for the function to be restored, the three solutions given by (7)–(9) are equivalent; expression (9) can be obtained from expression (8) [18]. Each method for solution is more or less computationally the same in that each requires $\sim O(N^3)$ multiplications; this was pointed out earlier in [1].

We note that expression (7) requires on the order of CN^3 multiplications, where the overhead C is large by comparison with the other methods. However, a primary concern is the ease with which the regularization parameter β can be varied; this can be done at the cost of $O(N^2)$ multiplications for (7).

V. Implementation on a Neural Network

We will now show how a superresolution algorithm equivalent to the previously described approaches can be defined on a Hopfield neural network. Several issues must be addressed. First, it is necessary to define the connection matrix from the cost function. For some problems this cannot be accomplished without performing more calculations than are required for a more conventional solution to the problem. The latter consideration was noted by Takeda and Goodman [19]. Thus, the computational load or complexity must be considered in deciding the merits of a neural network solution to this problem.

Other issues which must be addressed center on modifications of Hopfield's formulation to satisfy our requirements. Hopfield ensures that the network will converge by arbitrarily setting the diagonal of the connection matrix to zero. This is unacceptable, because it shifts the absolute energy minimum of the network from the minimum of (6). Moreover, while a suitable network can be constructed from two-state elements, one often requires that the reconstruction be represented over at least a set of grey levels. Thus, it may be necessary to combine a number of neural elements to represent a reconstruction pixel. The method of coding the grey levels depends on the complexity of the elements, i.e., whether they can only take on two states, a bounded continuum of values such as $0 \le v_k \le 1$ ("graded neurons"), or an (effectively) unbounded continuum. Granularity of the representation will affect the convergence properties of the network; coarsely quantized systems can converge to local energy minima, yielding less than optimal reconstructions.

We shall first demonstrate the formal mapping of a superresolution algorithm onto a Hopfield network. The technique will then be extended to fully address the second and third issues on a two-state network. We will briefly examine the advantages of more finely quantized systems, and finally discuss the relation between parallel thresholding and a regularized form of the well-known Gerchberg–Papoulis spectral extrapolation algorithm [20]–[22].

Let us represent the current estimate by the state of the network v. We can rewrite (6) as

$$E = v^TA^TAv - 2v^TA^Tg + g^Tg + \beta v^Tv.$$

Comparing this expression for E with that of the Hopfield network, (4), gives

$$\left.\begin{aligned} T &= -2(A^TA + \beta I) \\ b &= 2A^Tg \end{aligned}\right\} \tag{10}$$

where I is the identity matrix, and the g^Tg term can be ignored, since it represents a total offset for E.

Thus, superresolution can be mapped simply and directly onto a Hopfield network. The connection matrix is formed from the imaging operator matrix, which contributes information about the imaging system, and the regularization parameter β, which sets a bound on the norm of the final estimate. The available data g contribute only to the bias vector b.

For serial operation, the change in energy due to a change Δv_k is

$$\Delta E_k = -\Delta v_k[(Tv + b)_k + \tfrac{1}{2} T_{kk}\Delta v_k]. \tag{11}$$

Convergence to a minimum is guaranteed if the expression in braces always has the same sign as Δv_k. This can be ensured by altering the diagonal of T to zero; however, such a change is equivalent to choosing an arbitrary value for β. This is not acceptable, since the regularization parameter should be chosen to reflect the noise in the data and to obtain an optimum reconstruction, not to ensure convergence of the algorithm.

VI. Superresolution on a Two-State Network

In this section we will modify the Hopfield formulation so that the energy minimum of the network will coincide with that of (6), while still decreasing the energy with each change in the state vector. We shall find that this is possible by introducing a two-level threshold in place of the usual single-level one. In addition, we will incorporate a generalized grey scale mapping which describes linear transformations of a state vector v into a vector w having grey levels. We write this as $w = Sv$, where S could

be a mapping from an N element vector, each of whose elements can take the values 0 or 1, to an L-element vector ($L \leq N$) whose elements can take a wider range of values. For example, if S represents a base-2 mapping, each element of v can represent a power of 2, giving a range of $2^{N/L}$ values for each element of w. The range of v need not be limited to $\{0, 1\}$: we use this for the purpose of illustration. Other coding schemes are possible, such as clustering or bit-density coding.

The expression for the energy is now

$$E = \|Aw - g\|^2 + \beta\|w\|^2.$$

Suppose the grey-level vector w is perturbed by some amount Δw. The difference in energy between states is

$$\Delta E = 2\Delta w^T\{(A^TA + \beta I)w - A^Tg + \tfrac{1}{2}(A^TA + \beta I)\Delta w\} \tag{12}$$

or, in terms of the neural state vector v,

$$\Delta E = 2\Delta v^T[S^T(A^TA + \beta I)Sv - S^TA^Tg + \tfrac{1}{2}S^T(A^TA + \beta I)S\Delta v]. \tag{13}$$

It should be emphasized that (13) contains no assumption about the range of values of v or of the updating mode of the network. A restriction to two states reflects the desire to use a large number of simple binary processing elements in neural architectures.

We now present a procedure which ensures that the change in energy expressed by (12) and (13) always decreases, provided serial thresholding is adopted. For a change Δv_k, the change in energy ΔE_k of the network is given by (11), with the following definitions for T and b:

$$T = -2S^T(A^TA + \beta I)S$$

$$b = 2S^TA^Tg.$$

The grey-scale mappings we are considering associate a specific neuron with one and only one image pixel. Hence, the columns of S each contain only one element, and it is not difficult to show that the diagonal elements of T take the form

$$T_{kk} = -2S_{jk}^2(A^TA + \beta I)_{jj} \tag{14}$$

where S_{jk} is the nonzero element of the kth column of S. Since the diagonal elements of A^TA are positive, and β is some positive quantity, T_{kk} is always negative. Hence, we can rewrite (13) in the form

$$\Delta E_k = -\Delta v_k\{(Tv + b)_k - \tfrac{1}{2}|T_{kk}|\Delta v_k\}. \tag{15}$$

Thus, ΔE_k will be negative provided $(\Delta v_k)^2 < \Lambda_k \Delta v_k$, where

$$\Lambda_k = \frac{2}{|T_{kk}|}(Tv + b)_k$$

the maximum decrease in energy occurring when

$$\Delta v_k = \tfrac{1}{2}\Lambda_k. \tag{16}$$

Thus, we require that $|\Delta v_k| < |\Lambda_k|$ and $\operatorname{sgn}(\Delta v_k) = \operatorname{sgn}(\Lambda_k)$. For a binary network, where $v_k \in \{0, 1\}$, we then obtain the following rule to ensure that the network energy does not increase:

$$v_k^{(n+1)} = \begin{cases} 1 & \text{for } \Lambda_k > 1 \\ v_k^{(n)} & \text{for } |\Lambda_k| \leq 1 \\ 0 & \text{for } \Lambda_k < -1. \end{cases}$$

We consider next the operation of a nonbinary network.

VII. Superresolution on a Nonbinary Network

The restriction of the state vector v to binary values permitted the simplest possible processing elements to be used in a neural architecture. With more complex processors this simple representation is unnecessary and inefficient: the optimal coding scheme is intimately related to the nature of the available hardware.

A disadvantage of simple two-level elements is that they can give a coarsely quantized representation in reconstruction space which leads to the creation of local energy minima. There is still only one absolute energy minimum, but the network may converge to a local minimum of higher energy. It should be recognized that this behavior also occurs with a single level threshold and a zero-diagonal T matrix, and the resulting reconstructions are sometimes called "spurious stored states." A typical solution of this type is shown in Fig. 1(c) for a network of 90 two-level elements; a 6-b coding scheme yields 15 points in the reconstruction. Whether such a reconstruction is of acceptable quality is difficult to predict, and a function of the needs of the user. This difficulty can be overcome by using elements which can take on values over a continuum, such as $0 \leq v_k \leq 1$ (Fig. 1(e), (f)). These elements are similar to the graded neurons employed by Hopfield in a differential network [10].

We shall now examine the behavior of an asynchronous network composed of elements which can take on a continuum of values. Because $|\Delta v_k|$ is no longer fixed there is no need for a threshold/decision operator; we will simply use the value of Δv_k which yields the greatest decrease in energy.

It was noted above that the maximum decrease in energy occurs when

$$\Delta v_k = \Lambda_k/2.$$

In addition, $|\Lambda_k|$ represents an upper bound on $|\Delta v_k|$. As one approaches the solution, $(Tv + b)_k$ approaches zero, so this upper bound decreases. For networks with a fixed $|\Delta v_k|$ (e.g., ± 1 for all k), one would expect $|\Lambda_k|$ eventually to be smaller than $|\Delta v_k|$, so no changes can be made to reduce the energy, even though the network is not yet at the global minimum. Thus, the fixed step methods will generally be limited to some outer neighborhood of the global energy minimum (although one might arrive at the minimum).

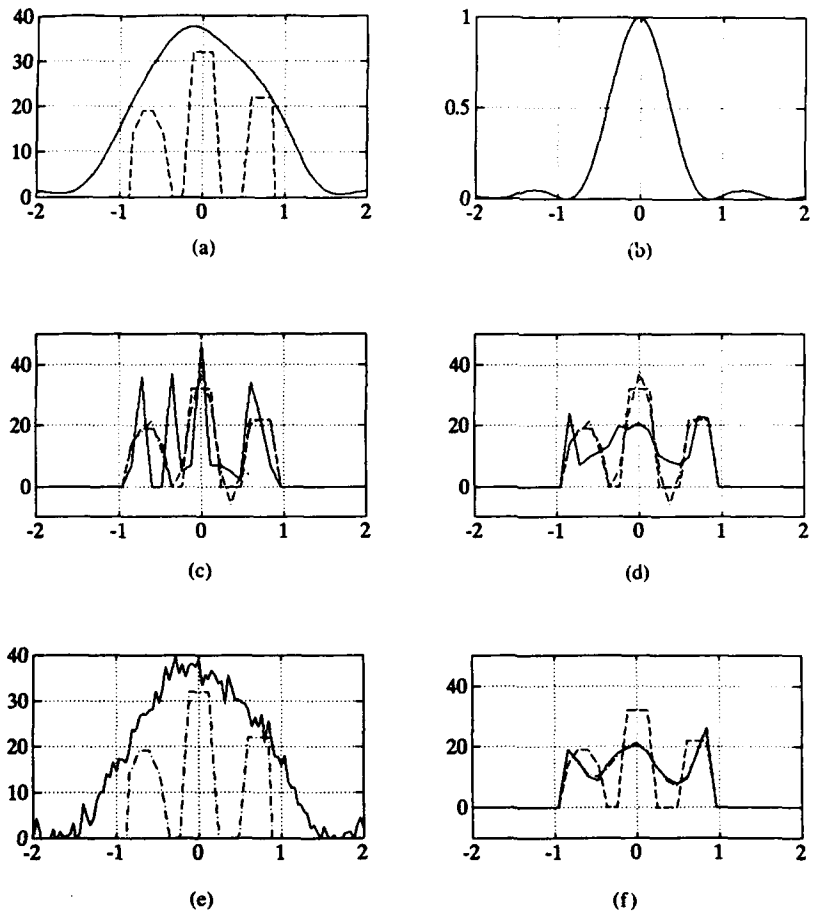


Fig. 1. The line types used are solid for an image or a neural network reconstruction; dashed for the object; and dot-dashed for the algebraic (SVD) reconstruction. (a) Object and incoherent image. Ninety two-state {0, 1} elements are mapped with a 6-b base 2 scheme to 15 pixels in the object. (b) The point-spread function of the imaging system (sinc^2). (c) Network using two-state neurons converges to a local energy minimum after 5 cycles, $\beta = 10^{-5}$. (d) After 5 cycles using 90 graded neurons the network has not converged, but a good estimate of the object has emerged, $\beta = 10^{-5}$. (e) Object and image containing 5% Gaussian additive noise. (f) Neural reconstruction from (e) after 50 cycles, $\beta = 10^{-4.5}$. Note that it is nearly indistinguishable from the SVD result.

However, if we adopt graded neurons it is still possible to use very simple elements, yet circumvent the finite step limitation. Since these elements can take on a continuum of values between two limits, the energy is forced to decrease at each step. A serially threshold network, constructed from these elements can therefore reach the global energy minimum after a sufficient number of cycles. One could also dispense with the coding scheme, and use a smaller number of more complex elements.

We would like to operate the network in the synchronous mode to make efficient use of the network's parallelism. If the kth neuron changes by Δw_k

$$w_k^{(n+1)} = w_k^{(n)} + \Delta w_k$$

where convergence is assured provided that the conditions of Section VI are met.

A regularized form of the Gerchberg–Papoulis algorithm reads [20]–[22]

$$w^{(n+1)} = A^T g + [(1 - \beta)I - A^T A]w^{(n)}$$
$$= w^{(n)} + (Tw^{(n)} + b).$$

Thus, if

$$\Delta w_k = (Tw^{(n)} + b)_k$$

parallel operation of the network will result in a computation which is identical to the regularized Gerchberg–Papoulis algorithm. Since the latter always converges [22], this choice for the Δw_k must always be possible. Optimal selection of the Δw_k to accelerate convergence of the network in the parallel mode is under investigation.

VIII. Computational Complexity of Neural Algorithm

The computational complexity associated with image reconstruction or superresolution using the singular value decomposition of A to solve (8), and using the neural network approach, has been examined for one-dimensional images (Table I). The computational load associated with the neural network is independent of whether the thresholding is serial or parallel, although the actual computational time is obviously less for parallel thresholding. The

TABLE I
THE NUMBER OF OPERATIONS REQUIRED FOR SUPERRESOLUTION BY SVD INVERSION AND BY A NEURAL OPTIMIZATION ARE LISTED FOR THE TOTAL CALCULATION, AND FOR UPDATING EITHER THE IMAGE OR THE REGULARIZATION PARAMETER

Requirement	Operation	SVD	Neural
New β	Mults	$N^2 + N$	$KN^2 + (K + 1)N$
	Adds	N^2	$KN^2 + (K + 1)N$
	Divs	N	N
New image	Mults	$2N^2 + N$	$(K + 1)N^2 + (K + 1)N$
	Adds	$2N^2$	$(K + 1)N^2 + KN$
	Divs	N	N
Total operations	Mults	$15N^3 + 3N^2 + 2N$	$N^3 + (K + 1)N^2 + (K + 1)N$
	Adds	N^2	$N^3 + KN^2 + KN$
	Divs	N	N

number of additions, multiplications, and divisions for each technique are listed in Table I for three situations. The first row gives the number of calculations needed for a new value of the regularization parameter β; the neural network has a disadvantage in this case because it must generally run for some K iterations. The second row gives the computational cost involved in updating the input image data vector g. The neural network once again is at somewhat of a disadvantage. However, by examining the total number of operations from the beginning, one can see that the neural approach is substantially more efficient because it calculates a matrix product once without the overhead associated with singular value decomposition; numerical experiments using a microcomputer indicate that the number of operations grows as $15N^3$ for an N point image. This is clearly increasingly significant for larger images.

IX. DISCUSSION AND CONCLUSIONS

Neural network solutions to image restoration problems are, therefore, competitive with, but not necessarily better than, more traditional methods for solving the problem (see also [22]). We have shown that both binary and nonbinary image reconstruction algorithms can be implemented on very similar neural architectures. The nonbinary case can be based upon network elements which take discrete (typically two-state) or continuous values. In the present application, the diagonal of the connection matrix is nonzero and, for the discrete element case, the crucial thresholding step was modified in order to ensure that the energy of the network decreases at each step; the thresholding step is unnecessary for the continuous case.

Thus, convergence to a minimum of the energy function is guaranteed for the both the discrete and continuous-element networks. However, this minimum is unique only in the continuous case, since discretization of the element states introduces local minima. Convergence has been demonstrated in this paper only for the case of serial thresholding; for parallel thresholding in the continuous case, the computation reduces to the regularized Gerchberg–Papoulis algorithm for a particular choice of the increments in the states of the network elements.

ACKNOWLEDGMENT

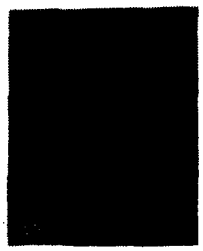
The authors acknowledge useful discussions with J. S. Bayley and C. L. Byrne.

REFERENCES

[1] J. B. Abbiss, J. S. Bayley, B. J. Brames, and M. A. Fiddy, "Superresolution and neural computing," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 880, pp. 100–106, Jan. 1988.
[2] J. A. Reggia and G. G. Sutton, "Self-processing networks and their biomedical implications," *Proc. IEEE*, vol. 76, pp. 680–692, 1988.
[3] J. A. Anderson and E. Rosenfeld, Eds., *Neurocomputing: Foundations of Research*. Cambridge, MA: M.I.T. Press, 1987.
[4] J. Y. Jau, Y. Fainman, and S. H. Lee, "Comparison of adaptive pattern recognition and image restoration with heteroassociative and autoassociative memories," in *Tech. Dig. O.S.A. Topical Meeting Opt. Comput.*, 1987, pp. 145–148.
[5] Y.-T. Zhou, R. Chellappa, A. Vaid, and B. K. Jenkins, "Image restoration using a neural network," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1141–1151, 1988.
[6] J.-S. Jang, S.-Y. Lee, and S.-Y. Shin, "An optimization network for matrix inversion," in *Neural Information Processing Systems*, D. Z. Anderson, Ed. AIP, 1988, pp. 397–401.
[7] B. Bai and N. H. Farhat, "Radar image reconstruction based on neural net models," in *Proc. IEEE APS/URSI Meeting* (Syracuse, NY), 1988, pp. 774–777.
[8] J. H. Winters, "Superresolution for ultrasonic imaging in air using neural networks," in *Proc. IEEE Int. Conf. Neural Networks* (San Diego, CA), 1988, pp. 609–611.
[9] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. U.S.*, vol. 78, pp. 2554–2558, 1982.
[10] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two state neurons," *Proc. Nat. Acad. Sci. U.S.*, vol. 81, pp. 3088–8092, 1984.
[11] J. Bruck and J. W. Goodman, "A generalized convergence theorem for neural networks and its applications in combinational optimization," in *Proc. IEEE Int. Conf. Neural Networks*, 1987, pp. 649–656.
[12] J. S. Bayley and M. A. Fiddy, "On the use of the Hopfield model for optical pattern recognition," *Opt. Commun.*, vol. 64, pp. 105–110, 1987.
[13] C. L. Byrne, R. M. Fitzgerald, M. A. Fiddy, A. M. Darling, and T. J. Hall, "Image restoration and enhancement," *J. Opt. Soc. Amer.*, vol. 73, pp. 1481–1487, 1983.
[14] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biolog. Cybern.*, vol. 52, pp. 141–152, 1985.
[15] N. H. Farhat and S. Miyahara, "Superresolution and signal recovery using models of neural networks," *OSA Top. Meeting Signal Recovery and Synthesis II*, 1986, pp. 120–123.
[16] R. Rastogi, P. K. Gupta, and R. Kumaresan, "Array signal processing with interconnected neuron-like elements," in *Proc. ICASSP*, 1987, pap. 54.8.1, pp. 2328–2331.

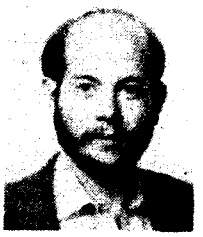
[17] G. Eichmann and M. Stojancic, "Superresolving signal and image restoration using a linear associative memory," *Appl. Opt.*, vol. 26, pp. 1911–1918, 1987.
[18] A. M. Darling, T. J. Hall, and M. A. Fiddy, "Stable, noniterative object reconstruction from incomplete data using *a priori* data," *J. Opt. Soc. Amer.*, vol. 73, pp. 1466–1469, 1983.
[19] M. Takeda and J. W. Goodman, "Neural networks for computation: Number representations and programming complexity," *Appl. Opt.*, vol. 25, pp. 3033–3046, 1986.
[20] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 735–742, 1975.
[21] R. W. Gerchberg, "Superresolution through error reduction," *Opt. Acta*, vol. 21, pp. 709–720, 1974.
[22] J. B. Abbiss, C. DeMol, and H. S. Dhadwal, "Regularized iterative and noniterative procedures for object restoration from experimental data," *Opt. Acta*, vol. 30, pp. 107–124, 1983.

*



John B. Abbiss received the B.Sc. degrees in mathematics and physics and the M.Sc. degree in solid state physics from the University of Wales, U.K., and later the Ph.D. degree in photon correlation techniques from the University of Surrey, U.K.

He spent the early part of his career working on guided missile development, after which he joined the Royal Aircraft Establishment at Farnborough. His research activities there involved probabilistic modeling for missile trials, determination of particle-size distributions from scattered-light measurements, and the development of photon correlation laser anemometry techniques for wind-tunnel applications. Other research interests included the development of improved methods for spectral estimation and band-limited signal extrapolation, and for image restoration from blurred and noisy data. Since joining Spectron (a division of the Titan Corporation) in California in 1985, he has been engaged in a wide range of projects, including the design and development of signal processing methods, software and instrumentation for laser velocimetry and optical spectroscopy in ground-based and airborne sensor systems, the further development of procedures for signal and image recovery, and investigations of applications of these techniques to satellite surveillance, together with their implementation in neural network form.



Bryan J. Brames was born in Ft. Wayne, IN. He received the bachelor's degree in physics from Purdue University, and the Ph.D. degree in optics from the University of Rochester in 1985.

He conducted postdoctoral work at the Blackett Laboratory, Imperial College. He then joined Spectron Development Laboratories in 1987, where he has participated in the development of optical air-data sensors, solid-state lasers, and signal processing, including phase retrieval and superresolution.

*

M. A. Fiddy (M'88) was born in Portsmouth, England, and attended the University of London. He received the first class B.Sc. degree in physics in 1973, and the Ph.D. degree in 1977.

He worked as a postdoctoral Research Assistant in the Physics Department at Queen Elizabeth College, University of London in 1977, and in the Department of Electronic and Electrical Engineering at the University College London in 1978. In 1979 he was appointed a Lecturer in Physics at Queen Elizabeth College and then moved to Kings College London in 1983. During part of 1982, he was a Visiting Associate Professor at the Institute of Optics. University of Rochester and in 1985–1986 he held a similar position in the Mathematics Department at the Catholic University of America, Washington, DC. In September 1987 he joined the faculty of the University of Lowell. He has worked for many years in the field of inverse problems and optical signal processing; he has written over 60 refereed publications and over 90 conference publications.

Dr. Fiddy is on the Editorial Board of the journals *Inverse Problems* (Institute of Physics), *Waves in Random Media* (Institute of Physics), *Multidimensional Systems and Signal Processing* (Kluwer), and *Optical Computing and Processing* (Taylor and Francis). He is also a member of OSA, IOP (U.K.), and SPIE.

PROCEEDINGS REPRINT

SPIE–The International Society for Optical Engineering

Reprinted from

Advanced Signal Processing Algorithms, Architectures, and Implementations II

24–26 July 1991
San Diego, California

Volume 1566

Restoration of sub-pixel detail using the regularized pseudo-inverse of the imaging operator

J. B. Abbiss and B. J. Brames

TITAN SPECTRON Division
3535 Hyland Avenue
Costa Mesa, California 92626

ABSTRACT

We present an analysis and computational results relating to the regularized restoration of subpixel information from undersampled data. The method makes use of a small set of images in various stages of defocus. An iterative implementation permits the incorporation of a non-negativity constraint. The problem we consider is fundamentally under-determined, but useful results can be obtained in reasonably low noise conditions.

1. INTRODUCTION

The investigations discussed here form part of a program whose subject is the enhancement of images obtained from space-based remote sensors. For the present purpose, these images are assumed to consist of quantized data from a fixed two-dimensional set of sensors, such as a CCD array. Typically for these arrays, the Airy disc is smaller than one pixel. Thus, for reasonably fast and well-corrected optics, the conventional limit on system resolution is likely to be the result of the spatial sampling associated with the pixel size; i.e., the integration of the light energy in the image over the area represented by each pixel.

For clarity, we consider the case of one-dimensional imaging with an incoherent source. Let the system point spread function (psf) be represented by the continuous imaging operator L . Then, for an isoplanatic system in the absence of noise, the image g of an object f is given by the convolution

$$g(y) = \int_S L(x - y)\, f(x)\, dx \tag{1}$$

where S is the support of f . The output of the k^{th} detector (pixel), extending from Y_k to Y_{k+1}, is

$$g_k = \int_{Y_k}^{Y_{k+1}} g(y)\, dy$$

$$= \int_{Y_k}^{Y_{k+1}} dy \int_S L(x - y)\, f(x)\, dx$$

Interchanging the order of integration (which is certainly permissible with the physical functions considered here) and making the definition

$$a_k(x) = \int_{Y_k}^{Y_{k+1}} L(x - y)\, dy \ ,$$

we obtain for the pixel-integrated image

$$g_k = \int_S a_k(x)\ f(x)\ dx,\ k = 1, 2, \cdots, K \tag{2}$$

We now discretize the object into a set $\{f_j, j = 1, 2 \cdots, J\}$ over equal intervals and write, as an approximation,

$$g_k = \sum_{j=1}^{J} a_{kj} f_j, \quad k = 1, 2, \cdots, K \tag{3}$$

where a_{kj} is the integral of $a_k(x)$ over the j^{th} interval. In matrix form, equation (3) can be written

$$g = Af$$

where A is a K x J matrix.

Since the a_{kj} can be determined from the system psf and the pixel array structure, the deconvolution problem represented by equation (3) can, in principle, be solved and the set $\{f_j\}$ reconstructed approximately from an equal set of measurements of the g_k.

In practice, the restoration problem is made much more difficult because of background and intrinsic noise, array imperfections and errors or uncertainties in the knowledge of the optical properties of the system. When pixel integration is over a significant part of the system psf, achieving even a modest degree of enhancement, using a single image, poses intractable difficulties in the presence of these perturbations. However, the necessary additional information can be derived from multiple differing images of the given object. After a brief discussion of the characteristics of the inverse problem represented by equation (3), we shall give a detailed description of a specific method for acquiring this information.

2. REGULARIZED IMAGE RESTORATION

There are many applications that require the restoration of a signal or image from a limited discrete data set; for example, samples of the spatial or temporal spectrum, or of the object's low or band-pass filtered image. An important *a priori* assumption in image restoration is that the object or objects are of compact support. This leads to the well-known result that their spectra are bandlimited functions. In principle, therefore, one might hope to extend limited spectral data by means of analytic continuation, and then, by Fourier transformation, obtain an enhanced image. This procedure is notoriously unstable in the presence of noise and does not provide a practical solution to the problem.

This central difficulty can be expressed in another way. Most signal or image recovery problems can be described by linear equations of the form

$$g(x) = \int A(x,y)\, f(y)\, dy$$

where A is the system point spread function or the Fourier transform kernel, for example. The interpretation of the data g(x) to obtain information about the object f(y) requires the solution of a linear inverse problem. This is equivalent to finding the solution of a Fredholm integral equation of the first kind. It is well-known in such a case that small fluctuations in the data g(x) can lead to very large fluctuations in the estimate of the unknown function f(y). This is a manifestation of the ill-posed nature of the problem (the inverse operator is unbounded) and some method of stabilization is needed to determine useful solutions.

The problem of a lack of continuous dependence on the data can be overcome by one of the various techniques of regularization [1, 2]. Essentially, the ill-posed problem is replaced by a related well-posed one, chosen to be physically meaningful and to possess the necessary properties of convergence and stability. Thus we may change the concept of a solution, or the Hilbert spaces of which the object and image are elements, or their topologies, or the operator itself. The technique we shall use belongs to the last category.

We impose physically reasonable constraints on the permitted solutions. If ε is a measure in the norm sense of the noise in the image (norm in the appropriate Hilbert space is denoted by $\| \cdot \|_H$) and if C is some constraint operator with E a known bound, we shall require that all possible reconstructions f' satisfy

$$||g - Af||_G \leq \varepsilon \quad \text{and} \quad ||Cf||_F \leq E$$

C may be used, for example, to impose smoothness on the reconstruction or to weight the reconstruction support. If C is the identity operator, E is a bound on the norm of the reconstruction. We combine the constraints quadratically and minimize the functional

$$||g - Af||_G^2 + \beta ||Cf||_F^2$$

where $\beta = \varepsilon^2/E^2$. Note that smaller values of β are equivalent to demanding greater fidelity between the reconstruction and the data; greater values place more emphasis on the property of the reconstruction associated with C. In the present discussion we shall take $C = I$. The minimizer f_β can then be expressed in either of the forms

$$f_\beta = (A^*A + \beta I)^{-1} A^* g$$

or (4)

$$f_\beta = A^* (AA^* + \beta I)^{-1} g$$

where A^* is the operator adjoint to A. The inverses of the bracketed operators will always exist, since the eigenvalues of the symmetric operators A^*A and AA^* are non-negative. The operator, the image and the reconstruction will consist in practice of finite arrays and equation (4) becomes a matrix equation.

3. SUB-PIXEL RESOLUTION FROM MULTIPLE DEFOCUSSED IMAGES

We can state the deconvolution problem in the more general case where we are given a set of R differing noisy images of the same object over the same pixel array.

Then the r^{th} image consists of the vector of pixel outputs given by the equation

$$g^{(r)} = A_r f + n^{(r)} \tag{5}$$

where $n^{(r)}$ represents some additive noise vector. (Other forms of noise can be accommodated by appropriate modifications of equations (4) and (5), but we shall not address this question here.) We are required to estimate f from the set $\{g^{(r)}, r = 1, 2, \cdots, R\}$.

We obtain the solution by assembling the image set into one composite image, and the corresponding matrices of integrated psf samples $a_{ij}^{(r)}$ into a single imaging matrix. The regularized solution is then again given by equation (4). The appropriate value for the regularization parameter β, which is closely related to the signal-to-noise ratio in the data, can be estimated in several ways; for example, by the method of weighted cross-validation [3].

The image set required to achieve sub-pixel resolution can be derived by various means. Stark and Oskoui [4] discuss an object reconstruction technique which uses a set of images differing from one another by rotation or lateral translation of the pixel array. It is evident that acquiring a sequence of data sets by lateral displacement of the detector (or image) through some fraction of a pixel at each step allows one to sample the image as finely as is desired. (For the two-dimensional image, the displacements can be in any direction.) Rotational displacement similarly permits arbitrarily fine sampling. They consider the specific case in which the system psf is much smaller than a pixel, so that resolution in the data is governed almost completely by pixel integration rather than the optical properties of the system.

An alternative method, which we shall adopt, is to use an image set obtained with differing point-spread functions. These could be generated by separate optical systems, or conceivably at different wavelengths. The system psf can also be conveniently altered by varying the degree of defocus; implementations of this method could depend on a single detector translated into the chosen planes of defocus, or a system of beamsplitters and detector arrays in appropriate locations.

If the reconstruction procedure is to be effective, the various images must contain significantly different information, which implies that the point spread functions must differ appreciably over the scale of a pixel. We shall assume that the images are formed on the same array, or arrays with identical characteristics, and that the object field is spatially and, if necessary,

temporally invariant from one image to another. We shall also assume that the point spread functions are accurately known and, for the purposes of the algorithm used later, that the images are formed from incoherent radiation. We do not require that any of the point spread functions are much smaller than a pixel; for the illustration presented below, the psf at focus is about half the size of a pixel. In a practical implementation, the images would not be centered on the pixel arrays in exactly the same way; i.e., there would be some lateral translation between the images. The effect of including lateral translations has not yet been investigated, but one would not expect reconstruction quality to suffer under these circumstances.

4. ILLUSTRATION

To illustrate these ideas, we include some results from a numerical simulation. The object field consists of a group of delta-function-like incoherent radiators. We shall show that from a small set of images, one at focus, the others at various stages of defocus, object locations can be well recovered, even when several of the objects' geometrical images lie within a single pixel. The algorithm is based on the regularized formula of equation (4); in addition, the calculation is iterated a small number of times to enforce non-negativity on the reconstruction.

The reconstruction is made initially into a spatial region defined by the central lobe of the focussed image, but using a finer grid. No prior assumptions are made about the locations or the number of objects within this region. For small, relatively isolated sources, some ringing will occur in the reconstruction, with associated negative pixel values. The support is progressively refined by eliminating these pixels at each iteration until an entirely positive reconstruction is obtained. The smallest object space which is consistent with the image data will yield the best reconstruction, and the problem, initially underdetermined, becomes finally an overdetermined one.

A modified form of this scheme, which would be appropriate for more extended objects, includes a weight matrix which biasses the next iteration against those pixels with negative values. The computation is significantly slower in this case, since the size of the reconstruction space remains constant. We also note the possibility of using a regularized form of the non-negative least-squares algorithm of Lawson and Hanson [5]. The relative performance of this procedure, also of course iterative, has not been fully evaluated.

In this example, the object field consists of eleven highly localized sources of equal intensity, distributed over a 3x3 block of image pixels; see Figure 1. (It should be noted that the reconstruction grid used did not coincide with the object grid; hence the objects cannot appear as single-pixel "points" in the reconstruction.) The central lobe of the system psf was about half the width of a pixel. Four independent images containing equal energy were generated over a 7x7 block of pixels, the first corresponding to a focussed system, the others at various stages of defocus. A method was devised for choosing defocus conditions with significantly different information content. Starting with the focussed image, with associated imaging matrix A_0, we wish to find a defocussed image whose imaging matrix A_1 is as independent as possible from A_0. The criterion used for this purpose was the magnitude of the condition number (the ratio of the largest to the smallest singular values) of the matrix formed from the center columns of A_0 and A_1. The range of defocus over which the search was made was from 0 to 4 waves. The combination selected was that possessing the smallest condition number. Knowing A_0 and A_1, A_2 was determined, and finally A_3. The merit functions (reciprocals of condition numbers) calculated for combinations of two, three and four defocus levels are shown in Figure 2. Note that a zero occurs whenever the variable degree of defocus coincides with that corresponding to one of the other imaging matrices; then the composite matrix is singular and its condition number becomes unbounded. The degrees of defocus chosen for the four images in this case were of magnitude 0, 0.8, 1 6 and 2.4 waves. The corresponding images are shown in Figures 3-6. The only readily-identifiable feature in the focussed image is that the central pixel is brighter than the surrounding ones.

Reconstructions were performed over the region defined by the central block of 3x3 pixels, with sampling seven times as fine as that in the data. Thus the reconstruction space initially consisted of 441 points, while the four images provided a total of 196 data points. The reconstructions obtained when each of the images was corrupted by additive Gaussian noise with standard deviation equal to 1% of the mean pixel content is shown in Figure 7. All of the objects are located close to their true subpixel positions. Figure 8 shows the result obtained when the noise level is increased to 5%. The reconstruction is still generally accurate, although there is now some distortion in object location and one or two small artifacts have begun to appear. A signal-to-noise ratio can be defined as the ratio for each image of the sum, on a pixel-by-pixel basis, of the signal power to the sum of the noise power. At the 5% level, this quantity varied between 33 and 28 dB. This example was designed

to be reasonably challenging, and spreading the objects further apart, or reducing their number, considerably increases the algorithm's robustness against noise.

5. CONCLUSIONS

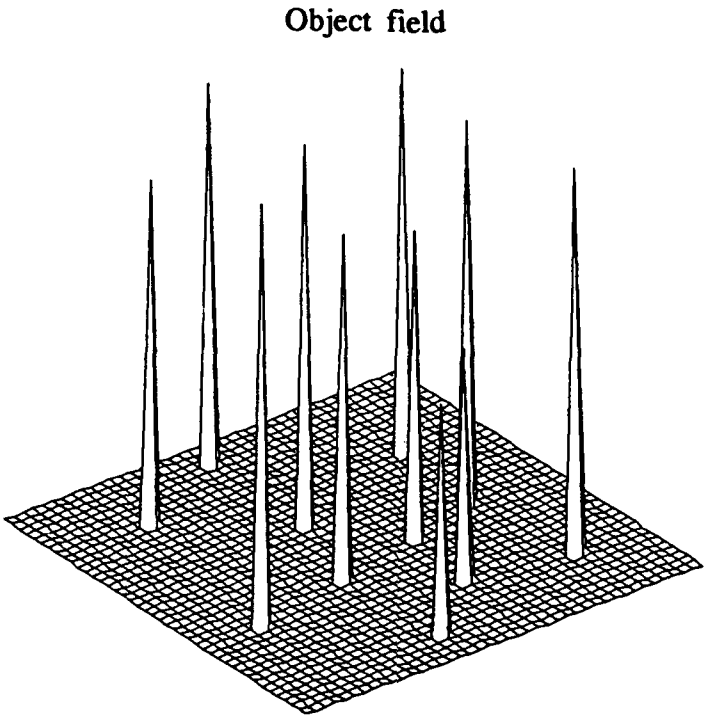
A method for recovering detail at the sub-pixel level from a small set of images in various stages of defocus has been discussed and demonstrated, using an algorithm which incorporates regularization to counter the destabilizing effects of noise. The iterative version of this algorithm, which permits the inclusion of a non-negativity constraint, is particularly effective at recovering accurate object support estimates. It is therefore an appropriate technique to use when it is known, *a priori*, that the object field consists largely of small well-separated targets. The sensitivity to noise of the method deserves more detailed investigation. A quantitative comparison of its performance with methods which make use of laterally translated or rotated image sets would also be of considerable interest. In practice there would inevitably be some lateral shift between images, even if they are formed on the same array, and a hybrid scheme might prove to be the most robust in the presence of noise. Ultimately, of course, the performance of any restoration algorithm must be limited by the information content of the image set. What should be sought, therefore, is the encoding scheme which most efficiently exploits the total information carried by the incident radiation.

6. ACKNOWLEDGEMENT

This work was supported by SDIO/IST and managed by ONR.

7. REFERENCES

1. A. N. Tikhonov, and V. Y. Arsenin, Solutions of Ill-Posed Problems, V. H. Winston and Sons, Washington, D.C. (1977).
2. K. Miller "Least squares methods for ill-posed problems with a prescribed bound," SIAM J. Math. Anal., 1, 1, 52-74 (1970).
3. G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," SIAM J. Numer. Anal. 14, 4, 651-667 (1977).
4. H. Stark, and P. Oskoui, "High-resolution image recovery from low resolution detector arrays," Digital Image Synthesis and Inverse Optics, Arthur F. Gnitro, Paul S. Idell, Ivan J. LaHaie, Editors, Proc. SPIE 1351, 80-84 (1990).
5. C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Prentice-Hall, New Jersey (1974).


Object field

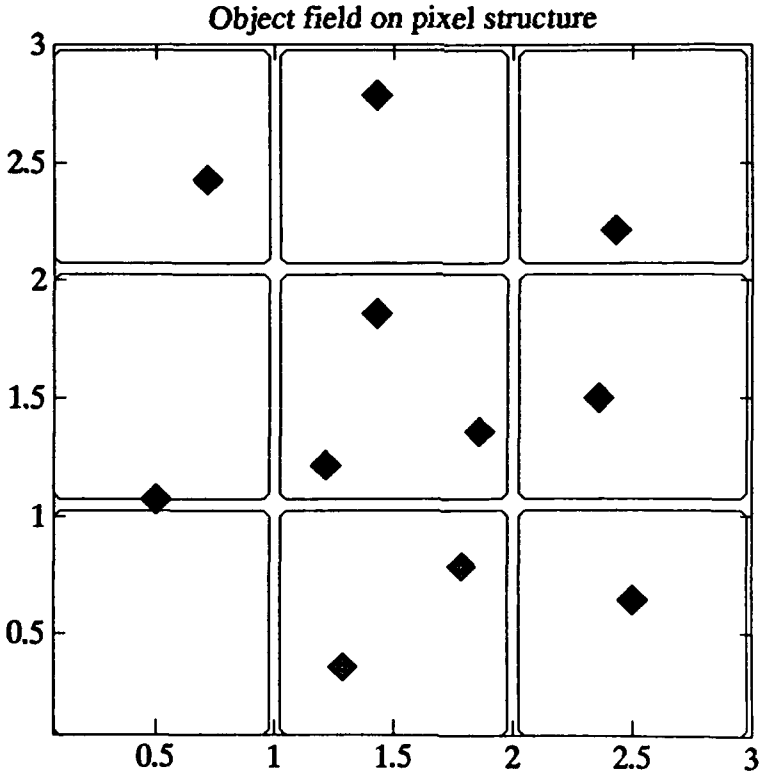

Object field on pixel structure
3
2.5
2
1.5
1
0.5
0.5
1
1.5
2
2.5
3

Fig. 1.

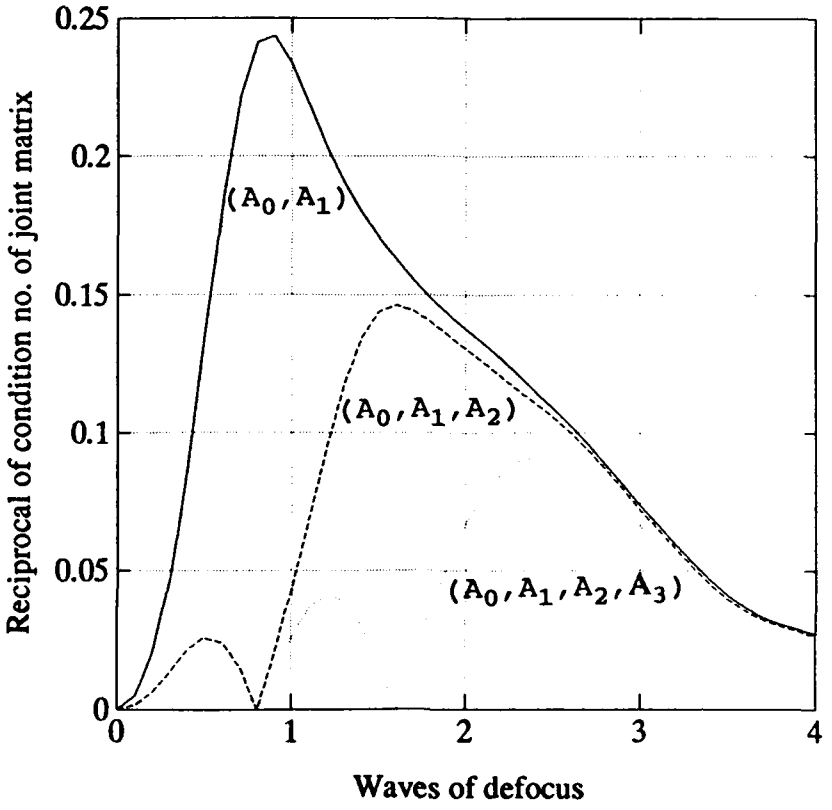


Fig. 2. Merit functions

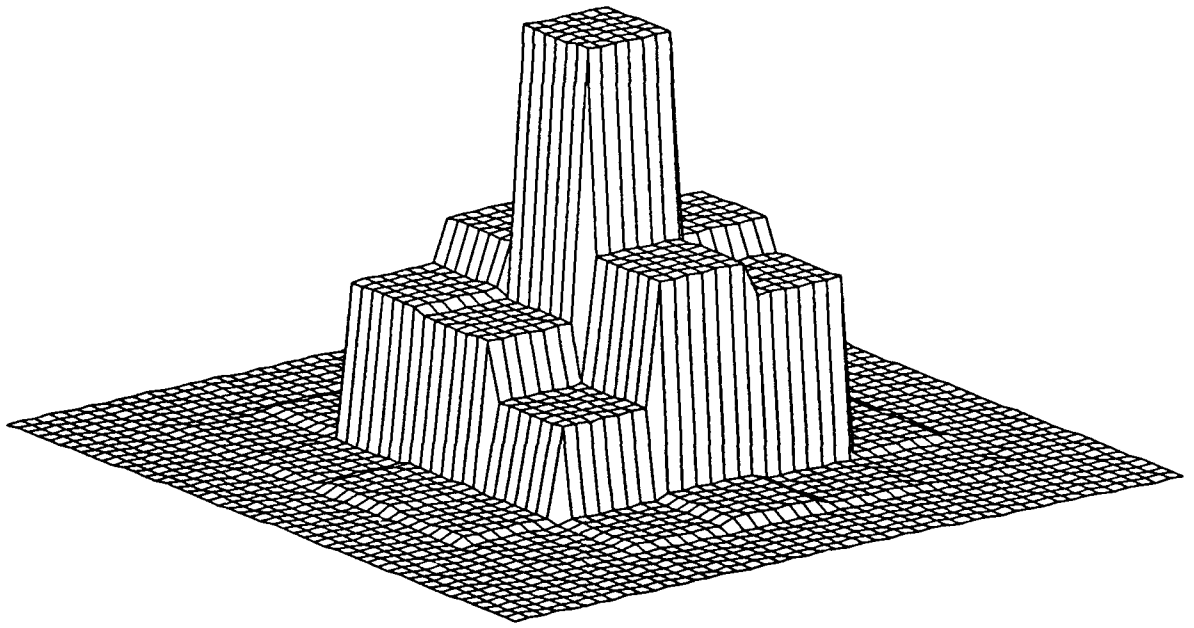


Fig. 3. Image at 0.0 waves defocus

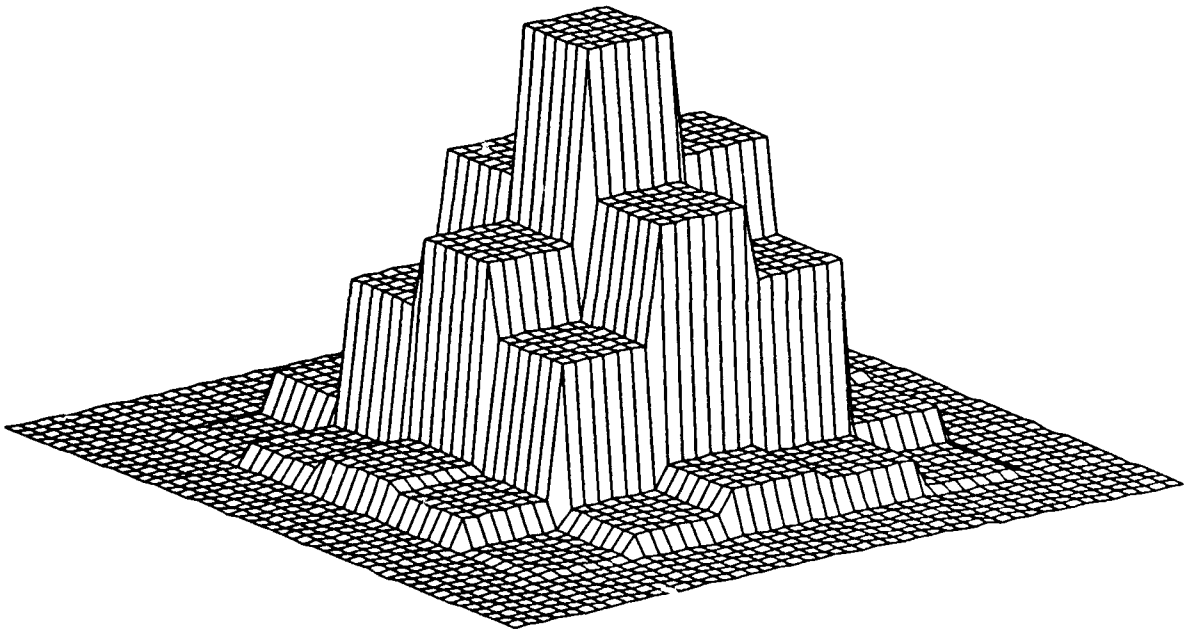


Fig. 4. Image at 0.8 waves defocus

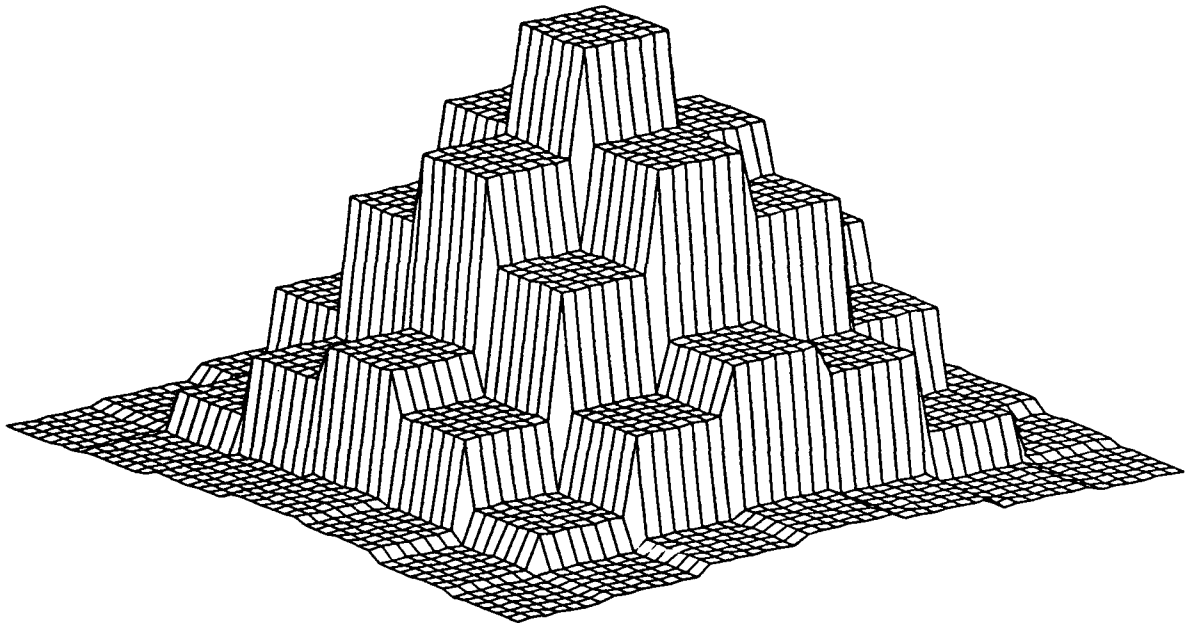


Fig. 5. Image at 1.6 waves defocus

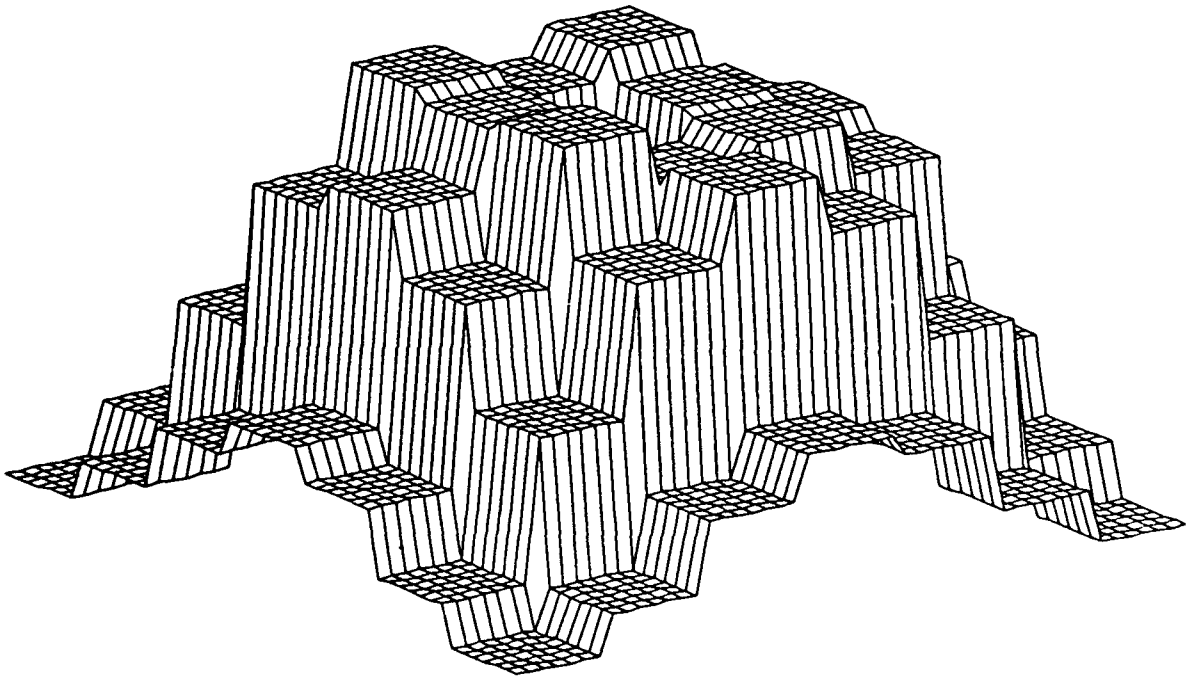


Fig. 6. Image at 2.4 waves defocus

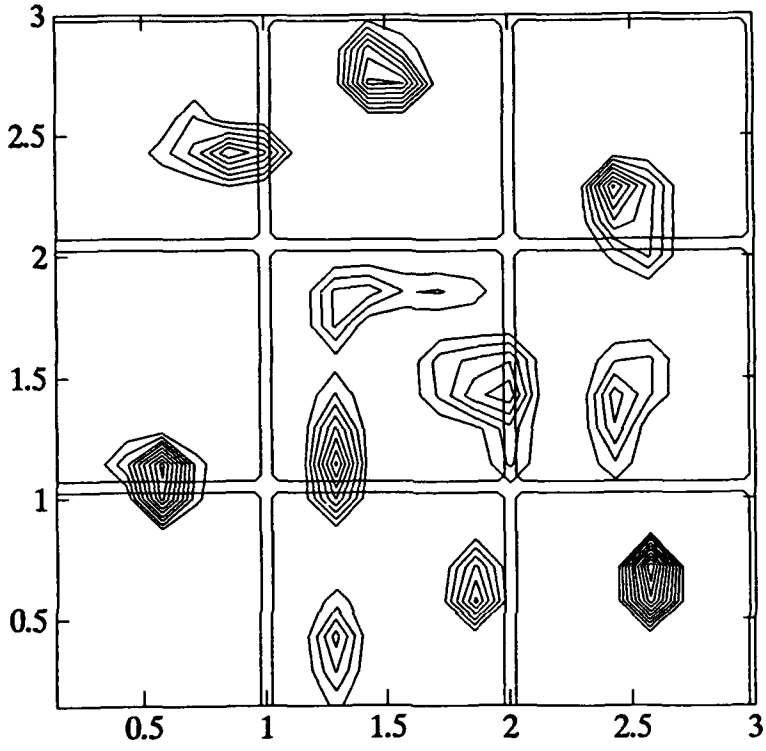


Fig. 7. Reconstruction from image with 1% added noise (beta = 5×10^{-5})

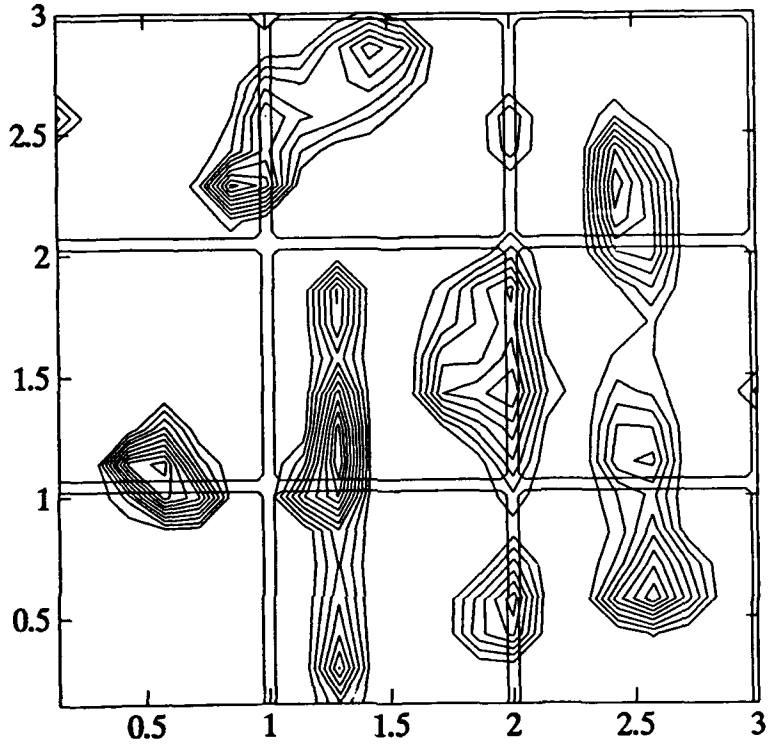


Fig. 8. Reconstruction from image with 5% added noise (beta = 10^{-3})

FAST IMAGE RECONSTRUCTION BASED ON THE REGULARIZED PSEUDOINVERSE OF THE IMAGING OPERATOR

John B. Abbiss

TITAN Spectron
1582 Parkway Loop
Suite B
Tustin, CA 92680-6505

Jeffery C. Allen, Richard P. Bocker, Harper J. Whitehouse

Naval Command, Control and Ocean Surveillance Center
RDT&E Division
San Diego, CA 92152-5000

ABSTRACT

A new algorithm for the restoration of extended images, the Regularized Pseudoinverse Deconvolution (RAPID) algorithm, is proposed. The algorithm consists of expanding the regularized pseudoinverse of the imaging operator into a sequence of terms which can be easily implemented using Fourier processing techniques. The first term of the expansion is closely related to generalized Wiener filtering if the point spread function is shift-invariant. The other terms in the expansion are correction terms which are small when the point spread function is shift-invariant, as is the case with many imaging systems. Even when the point spread function of the imaging system is space-variant, such as with a partially obscured imaging system or a system with severe aberrations, the correction terms are both few in number and easily implemented.

1. INTRODUCTION

In the absence of any other degrading effects, the performance of an optical system is ultimately restricted by diffraction. The finite extent of the entrance pupil imposes a fundamental upper limit on the system's spatial frequency response. The image quality of most operational systems will not, however, approach this theoretical limit very closely. It is possible that the design or construction will be flawed, as in the case of the Hubble Space Telescope, through defective manufacture, assembly or quality assurance procedures. In addition, aging of components will almost certainly compromise sensor performance at some stage in its lifetime, and, as in spaceborne operation, replacement may not be a simple task. The detector itself may impose limitations; for example, where a CCD array is used, information is lost in the inter-pixel areas, and image energy is integrated over the active area of each pixel. Other degrading factors will include defective pixels, noise in the CCD array and electronic subsystems, and a possibly obtrusive background. The methods of image restoration considered here were originally aimed at achieving performance beyond the diffraction limit [1], but are in fact capable of compensating simultaneously or separately for aberrations induced by the optical components and for the limitations of the detector. They are inherently robust and possess valuable noise-suppressing properties.

0-8194-0940-5/92/$4.00

The assumption is made that the overall effect of the optics can be described as a possibly time- and shift-variant blurring of the image due to diffraction and aberrations. Thus, at any instant of time, the point spread function may change across the sensor field-of-view. It can be assumed, however, that at any given point in the image the point spread function is effectively determined by its spatial location and is time-invariant over the integration time of the sensor and then undergoes a change at a later time. This is the case with the Hubble Space Telescope. It will be assumed that the set of point spread functions is known or can be measured. For the Hubble wide-field planetary camera there are two primary components to the point spread function; one due to diffraction by the aperture obstructions and the other due to spherical aberration. The point spread function is observed to be locally shift-invariant, and the image can be considered to be created by the summation across the entire field-of-view of the segmented set of localized point spread functions convolved with objects in the corresponding parts of the field.

However, for a simple refracting telescope designed for observation of extended objects, the aberrations causing image degradation can be highly space-variant. For an extended object, image segmentation introduces undesirable edge effects at the block boundaries when locally shift-invariant approximations of the globally shift-variant point spread function are used to process the separate blocks. In addition, post processing interpolation or iteration is required to smooth the block boundaries in the final composite image [2]. In the approach proposed in this paper, the point spread function is allowed to vary continuously over the whole image and a special decomposition of the image reconstruction operator is performed which permits Fourier techniques to be used to process the entire image.

The overall optical system imaging equation can be written as a Fredholm integral equation of the first kind. Solutions to ill-posed problems of this type are known to be numerically unstable [3, 4]. Additionally, it is anticipated that the image will be spatially sampled by a solid-state sensor which will introduce spatial integration, discretization and associated noise processes. Thus, after scanning the image into a vector, the integral representing the continuous image can be rewritten as a matrix expression. In general, the presence of the sensor noise takes the measured image vector out of the span of the columns of the kernel matrix, which is typically highly ill-conditioned. Thus, when it is desired to estimate what the image would have been in the absence of aberrations or with less diffraction than the instrument can provide, techniques derived from regularization theory are required to restore stability to the reconstruction. By introducing a suitable error criterion (based on, *e.g.*, a vector norm), images can be constructed which are, in terms of the chosen criterion, closer to the undistorted geometrical image of the object than the detected image data.

To the extent permitted by the noise in the image, in-band effects can usually be removed by some form of pseudoinverse filter [5]. However, detector pixellation and the finite aperture of any system set resolution limits not so easily overcome, and a method for achieving spectral extrapolation has to be devised. The spatial spectrum of the object is the Fourier transform of its amplitude, in the coherent case, or its intensity, in the incoherent case. If the object is known to be of finite extent, its Fourier transform is an analytic function, and the out-of-band part of the spectrum can in principle be fully recovered by analytic continuation [6] of the image spectrum after removal of any in-band distortion. The inverse Fourier transform of this extended spectrum would then yield a perfect image of the original object. Equivalently, one could attempt to solve directly the equation describing the imaging process. This, however, involves the inversion of an ill-conditioned matrix, and the restoration process is intrinsically unstable, even small amounts of noise rendering the results meaningless. These difficulties may be surmounted by applying the methods of regularization theory [7], developed to deal with ill-posed problems of this type; the solution is derived by means of a constrained least-squares procedure in which a regularization parameter plays an essential role. Stability in the restored image, which is computed via the regularized pseudoinverse of the imaging matrix, is controlled by this parameter. Its optimal value depends on the signal-to-noise ratio in the data.

2. NATURE OF THE PROBLEM

We wish to estimate the object f from an image g, given that

$$g = Af + r \tag{1}$$

where A is the imaging operator and r represents the corrupting effect of additive noise. For clarity in the analysis, we consider the one-dimensional case with the operator A given in integral form by

$$(Af)(y) = \int_a^b A(x,y) f(x)\, dx, \qquad c \le y \le d. \tag{2}$$

In the absence of noise, Eq.(1) becomes a Fredholm equation of the first kind, in which the unknown function appears only under the integral sign. We can identify the sources of difficulty in solving this equation in the presence of noise or other perturbations (such as computer round-off error) by means of a singular function analysis [8].

We expand the kernel of the integral in terms of the singular functions $u_i(x)$ and $v_i(y)$, orthonormal systems in object and image spaces respectively, and the singular values σ_i:

$$A(x,y) = \sum_{i=1}^{\infty} \sigma_i\, u_i(x)\, v_i(y). \tag{3}$$

The object and image can be expanded in the forms

$$f(x) = \sum_{i=1}^{\infty} f_i\, u_i(x) \tag{4}$$

$$g(y) = \sum_{i=1}^{\infty} g_i\, v_i(y) \tag{5}$$

where the coefficients are related to $f(x)$ and $g(y)$ by the integral formulae

$$f_i = \int_a^b f(x)\, u_i(x)\, dx \tag{6}$$

and

$$g_i = \int_c^d g(y)\, v_i(y)\, dy. \tag{7}$$

In the noiseless case $(r = 0)$, we find from Eqs. (2) and (6)

$$(Af)(y) = \sum_{i=1}^{\infty} \sigma_i\, f_i\, v_i(y). \tag{8}$$

It follows, using Eqs.(5) and (8), that

$$g_i = \sigma_i f_i \tag{9}$$

and hence

$$f(x) = \sum_{i=1}^{\infty} \frac{g_i}{\sigma_i} u_i(x). \tag{10}$$

Thus the object function can in principle be perfectly reconstructed from the set $\{g_i\}$ of image coefficients.

Now consider the effects of noise. By expanding $r(y)$ in terms of the $v_i(y)$, we can derive the contribution of the noise to the new image coefficients:

$$g'_i = \sigma_i f_i + r_i. \tag{11}$$

The estimate of $\hat{f}(x)$ is now

$$\hat{f}(x) = f(x) + \sum_{i=1}^{\infty} \frac{r_i}{\sigma_i} u_i(x). \tag{12}$$

Image formation is characteristically described by an integral transform of convolution type, *i.e.*, A is a convolution operator. Its singular-value spectrum typically decays asymptotically at an exponential rate [9]. Since the r_i will in general decrease less quickly, the sum in Eq.(12) will be divergent and no bound will exist for the 'distance' (in the sense of some appropriate metric) between the true object and the reconstruction. The effect of the noise on the reconstructed image is a manifestation of the fact that convolution is a strongly smoothing process - closely similar images can correspond to widely differing objects. Thus image restoration is an ill-posed problem, small perturbations in the data causing large changes in the solution represented by Eq.(12).

3. THE REGULARIZED SOLUTION

The methods of regularization theory [7, 10] can be exploited to convert the problem to a related well-posed one, *i.e.*, one for which the solution exists, is unique and depends continuously on the data. Since we shall later be concerned with computed reconstructions, we henceforth consider the problem in its finite discretized form; thus the imaging operator A becomes a matrix. Although strictly we should introduce new symbols, for convenience we continue to use f, g, and A to denote the discrete forms of object, image, and imaging operator. Generally $f \in C^n$, $g \in C^m$ and $A \in C^{m \times n}$.

To regularize the problem, we shall modify A. We impose constraints [11] on possible solutions f' by requiring that

$$\| Af' - g \|^2 \leq \varepsilon^2 \tag{13}$$

where ε is some suitable measure of the noise in the image, and that

$$\| f' \|^2 \leq \xi^2 \tag{14}$$

where ξ is some suitable measure of the permitted 'signal strength' of the solution. ($\| \bullet \|$ denotes norm in the Hilbert spaces associated with object and image.) We combine these constraints and minimize

$$\| Af' - g \|^2 + \beta \| f' \|^2$$

where the regularization parameter β is given by

$$\beta = \varepsilon^2 / \xi^2 \quad . \tag{15}$$

The minimum-norm solution to this constrained least-squares problem is given by

$$f_\beta = A_\beta^+ g \tag{16}$$

where

$$A_\beta^+ = (A^H A + \beta I)^{-1} A^H. \tag{17}$$

We note the relationship of A_β^+ (which we shall call the regularized pseudoinverse) to A^+, the Moore-Penrose pseudoinverse [12]

$$A^+ = \lim_{\beta \to 0} A_\beta^+. \tag{18}$$

The inverse of $(A^H A + \beta I)$ always exists, since $A^H A$ is non-negative definite and the regularization parameter β is positive. A value of β should be chosen which balances data fidelity against smoothness in the reconstruction. Methods are also available for determining β from the image data themselves [13, 14].

4. SINGULAR VALUE AND FOURIER DECOMPOSITIONS

It is often convenient to compute the regularized pseudoinverse via the singular value decomposition (SVD) of A:

$$A = U \Sigma V^H \tag{19}$$

where [15]

$$U^H U = I_m, \qquad V^H V = V V^H = I_n \tag{20}$$

and

$$\Sigma = diag\,(\sigma_1, \sigma_2, \ldots, \sigma_n), \qquad \sigma_i \geq 0. \tag{21}$$

The singular values $\{\sigma_i\}$ are assumed to have been arranged in descending order of magnitude

$$\sigma_1 > \sigma_2 > \sigma_3 > \ldots > \sigma_n. \tag{22}$$

Then we find

$$f_\beta = V \; \Sigma_\beta^+ \; U^H \; g \tag{23}$$

where

$$\Sigma_\beta^+ = diag\left(\ldots, \frac{\sigma_i}{\sigma_i^2+\beta}, \ldots\right). \tag{24}$$

This representation is useful when the behavior of the reconstruction as a function of the regularization parameter is being studied. Regularization can equivalently be achieved by setting β to zero, and simply truncating the singular value series at a point which is dependent on the noise level [16].

An *ab initio* computation of f_β via Eq.(23) requires the SVD of A followed by two matrix-vector multiplications. If the regularized pseudoinverse can be precomputed, only one matrix-vector product is needed to generate the reconstruction. For images of more than modest sizes, however, the computation rapidly becomes burdensome. If, for example, f and g are 100x100, A is a 10^4-by-10^4 matrix, and the matrix-vector product requires 10^8 multiplications. There will also be considerable storage demands. Thus, if major computational resources are not available, some means of simplifying the calculation will be needed in many cases of practical interest.

If the matrix A were square circulant of order n ($A = [a_{j-i+1}]$, subscript mod n), we could dramatically reduce the computational burden by exploiting the fact that the Fourier transform diagonalizes a circulant [17]. If F denotes the Fourier matrix:

$$F^H = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & w & w^2 & \ldots & w^{n-1} \\ 1 & w^2 & w^4 & \ldots & w^{2(n-1)} \\ & & & \ldots & \\ 1 & w^{n-1} & w^{2(n-1)} & \ldots & w^{(n-1)^2} \end{bmatrix}, \qquad w = exp(i2\pi/n) \tag{25}$$

then

$$A = F^H \Lambda F \tag{26}$$

where

$$\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n). \tag{27}$$

It follows from Eq.(16) that

$$f_\beta = F^H \Lambda_\beta^+ F g \tag{28}$$

where

$$\Lambda_\beta^+ = diag\left(\ldots, \frac{\bar{\lambda}_i}{|\lambda_i|^2+\beta}, \ldots\right) \tag{29}$$

and $\bar{\lambda}_i$ is the complex conjugate of λ_i. We note that the singular values $\{\sigma_i\}$ and the eigenvalues $\{\lambda_i\}$ are related by

$$\sigma_i = |\lambda_i|. \tag{30}$$

The operational counts for the SVD and the FFT are $O(n^3)$ and $O(nlogn)$, respectively.

5. STRUCTURE OF THE IMAGING MATRIX

Under some circumstances the imaging matrix can be readily modified to circulant form. For a shift-invariant one-dimensional system, the image is a convolution of the point spread function and the object. If the sampling intervals in image and reconstruction spaces are equal, the matrix A is then Toeplitz ($A = [a_{j-i}]$). If, in addition, the image and reconstruction vectors are of equal length, A can be padded to circulant form. In two dimensions, f and g are matrices and must be mapped into vectors. The manner in which this is performed will determine the structure of A. For column-wise mapping, for instance, again with equal sampling in image and reconstruction spaces, A becomes block-Toeplitz with Toeplitz blocks. If the image and reconstruction matrices have the same number of elements, A can be padded to become block-circulant with circulant blocks, and the problem is again amenable to Fourier transform methods. In both one and two dimensions, it should be noted that the penalty associated with the expansion of A to circulant or block-circulant form is a relaxation of the support constraint on the reconstruction, which renders the calculation, and in particular the degree of resolution enhancement achieved, much more sensitive to noise [18]. An alternative construction in the two-dimensional case is to zero-pad f and g into larger matrices and then to use a linear congruential scan to map the padded matrices into vectors. A then becomes a circulant matrix since the linear congruential scan is an isomorphism between 2D convolution and 1D convolution [19].

For less structured imaging matrices (*e.g.*, if the system is weakly space-variant) it may be asked whether accelerated computation of the matrix-vector product is still possible. In this context, recent work on circulant approximations to matrices of quite general form [20] appears highly relevant, and includes the following result. For any matrix, A_β^+ say, we can write

$$A_\beta^+ = C_o + \sum_{m=1}^{\alpha} L(x_m)\, C^T(y_m) \tag{31}$$

where C_o is a circulant matrix with the same last row as A_β^+, $L(x_m)$ is a lower triangular Toeplitz matrix with x_m as its first column, and $C(y_m)$ is a circulant matrix whose last row is y_m. The $\{x_m\}$ and $\{y_m\}$ may be obtained from the truncated SVD of the cyclic displacement of A_β^+:

$$A_\beta^+ - EA_\beta^+E^T = \sum_{m=1}^{\alpha} x_m y_m^T \tag{32}$$

where E is the cyclic downshift matrix

$$E = \begin{bmatrix} 0\,0\,0 \ldots 0\,1 \\ 1\,0\,0 \ldots 0\,0 \\ 0\,1\,0 \ldots 0\,0 \\ \quad\quad \ldots \\ 0\,0\,0 \ldots 1\,0 \end{bmatrix}. \tag{33}$$

For imaging matrices with strongly Toeplitz features, α should be a small number.

6. THE REGULARIZED PSEUDOINVERSE DECONVOLUTION ALGORITHM

Consider a two-dimensional optical imaging system whose point spread function is both time- and space-invariant. In the presence of additive noise, the imaging equation connecting the input (extended object), the output (degraded image), and the point spread function (impulse response) is given by the following two-dimensional convolutional integral equation

$$i(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} o(x',y')\,p(x-x',y-y')\,dx'dy' + n(x,y). \tag{34}$$

In shorthand notation

$$i = p\bigstar\bigstar o + n \tag{35}$$

where $o(x,y)$ represents the extended object, $i(x,y)$ the degraded image, $p(x,y)$ the point spread function, and $n(x,y)$ the noise. It is assumed that the noise is independent of position in the image.

The discrete version of Eq.(34) can, of course, be cast[21] into the following vector-matrix form

$$g = Af + r. \tag{36}$$

When, in particular, f, g, and r represent the one-dimensional column vectors formed by stacking the rows or columns of the discretized versions of the input $o(x,y)$, output $i(x,y)$, and the noise $n(x,y)$, respectively, the two-dimensional imaging matrix A is block-Toeplitz with Toeplitz blocks. It can be shown, using known results [22] that the regularized pseudoinverse A_β^+ is block-persymmetric with persymmetric blocks. The inverse of a Toeplitz matrix is persymmetric, and persymmetric matrices obtained by inverting Toeplitz matrices have much more Toeplitz-like structure than general persymmetric matrices. In particular, their displacement rank is the same as that of the parent-Toeplitz matrix[23, 24]. Displacement rank, it should be noted, is a quantitative measure of the closeness of a given matrix to being Toeplitz. In one dimension, the displacement rank of a Toeplitz matrix is ≤ 2, and the displacement rank of A_β^+ is ≤ 4.

In the early stages of this investigation, one of the authors (R.P.B.) noted that for a variety of point spread functions being studied the corresponding computer-generated regularized pseudoinverse matrices appeared to have banded block-Toeplitz structure. He considered the implication of this observation. In particular, if a space-invariant point spread function used in a two-dimensional linear convolutional imaging equation leads to a block-Toeplitz imaging matrix with Toeplitz blocks, then the converse must also be true. That is, given a block-Toeplitz imaging matrix containing Toeplitz blocks, then the corresponding space-invariant point spread function which gave rise to this imaging matrix could be easily ascertained. This implies that from the regularized pseudoinverse an inverse point spread function $d_\beta(x,y)$ could be constructed which could be used to process the image $i(x,y)$ and form an estimate $\hat{o}_\beta(x,y)$ of the original object. This two-dimensional linear convolution technique is summarized by the equation

$$\hat{o}_\beta = d_\beta\bigstar\bigstar i. \tag{37}$$

The technique was tested on a digital computer with encouraging results. Further experimental investigations indicate that results obtained with this Regularized Pseudoinverse Deconvolution (RAPID) algorithm are comparable in quality to those obtained using parametric Wiener filtering. An error analysis is given in Appendix A. Results using degraded images processed with both the RAPID algorithm and parametric Wiener filtering are presented in section 8.

7. DECONVOLUTION VIA WIENER FILTERING

Wiener filtering [21] is a well-known technique for processing images degraded and corrupted by noise as described by Eq.(34). From a knowledge of the point spread function $p(x,y)$ characterizing the optical imaging system, it is possible to compute the corresponding optical transfer function $\tilde{p}(f_x, f_y)$ using the two-dimensional Fourier transform

$$\tilde{p}(f_x, f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\, exp\,[-i2\pi(xf_x + yf_y)]\, dxdy. \tag{38}$$

From the optical transfer function and knowledge of the power spectra of object $S_o(f_x, f_y)$ and noise $S_n(f_x, f_y)$, the parametric Wiener filter can be constructed, namely

$$\tilde{w}_\gamma(f_x, f_y) = \frac{\left|\tilde{p}(f_x, f_y)\right|^2}{\tilde{p}(f_x, f_y)\left[\left|\tilde{p}(f_x, f_y)\right|^2 + \gamma\left|\frac{S_n(f_x, f_y)}{S_o(f_x, f_y)}\right|\right]}. \tag{39}$$

When $\gamma = 1$, Eq.(39) reduces simply to the Wiener filter. If γ is variable we refer to this as the parametric Wiener filter. In the absence of noise, either form of the Wiener filter reduces to the ideal inverse filter. When the power spectra are not known, which is often the case in practice, Eq.(39) can be approximated by

$$\tilde{w}_\gamma(f_x, f_y) \approx \frac{\left|\tilde{p}(f_x, f_y)\right|^2}{\tilde{p}(f_x, f_y)\left[\left|\tilde{p}(f_x, f_y)\right|^2 + \gamma\right]}. \tag{40}$$

From the Wiener filter, using either Eqs.(39) or (40), an inverse point spread function $w_\gamma(x,y)$ can be constructed using the two-dimensional inverse Fourier transform. That is,

$$w_\gamma(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{w}_\gamma(f_x, f_y)\, exp\,[+i2\pi(xf_x + yf_y)]\, df_x df_y. \tag{41}$$

With the inverse point spread function given by Eq.(41), an estimate $\hat{o}_\gamma(x,y)$ of the object can be computed using the two-dimensional linear convolutional equation

$$\hat{o}_\gamma = w_\gamma \star\star i \tag{42}$$

where, again, $i(x,y)$ is the degraded image.

8. RESULTS

Preliminary results obtained using the RAPID algorithm are presented in this section. For purposes of comparison, results obtained using the parametric Wiener filter algorithm are also included. The optical system considered for these studies was a simple spherical converging lens as shown in Fig. 1.

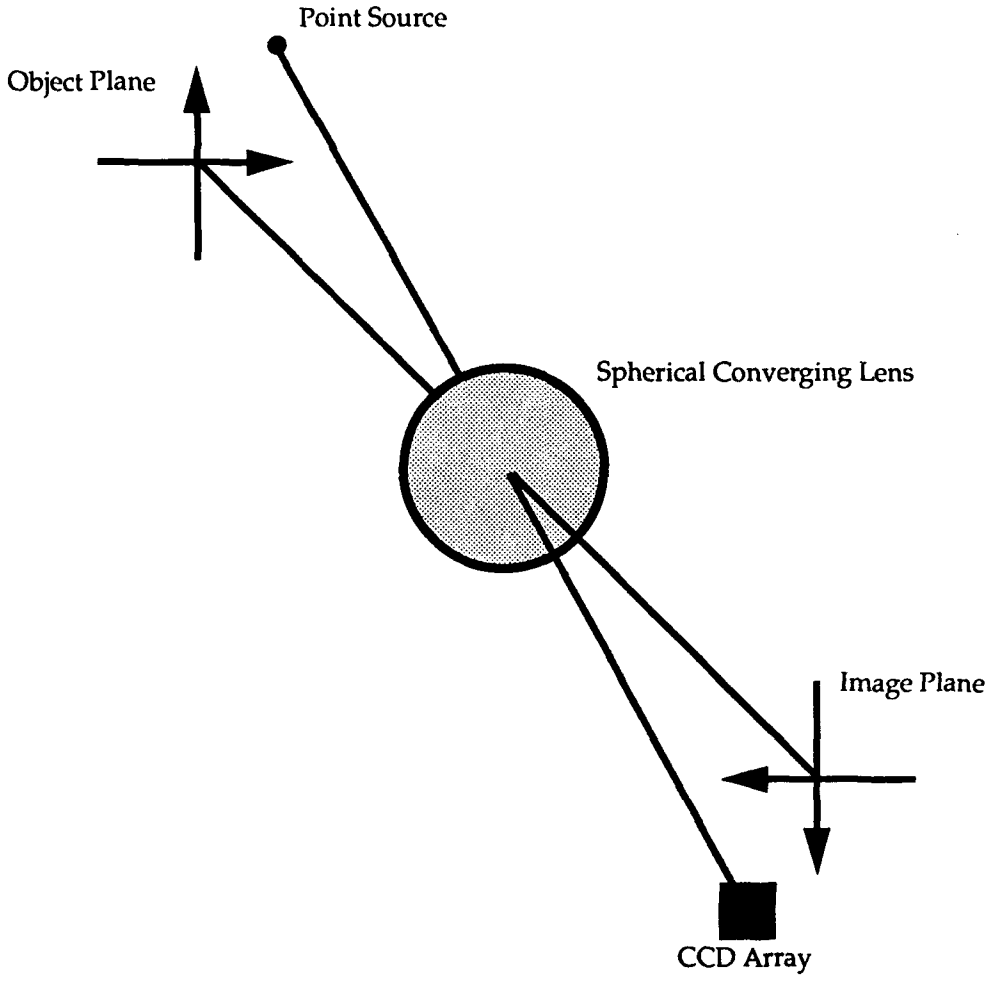


Figure 1. A Simple Spherical Converging Lens Imaging System

The object and image planes are coplanar and orthogonal to the optical axis of the lens. An arbitrary point source located in the object plane gives rise to an intensity distribution (the point spread function) in the image plane. A charge-coupled device (CCD), for example, can be used to measure the point spread function. A ray-trace program was used in the synthesis of four different point spread functions dominated by spherical aberration (Fig. 2), coma (Fig. 3), astigmatism (Fig. 4), and defocus (Fig. 5). The object plane distance (mm), image plane distance (mm), focal length (mm), F-number, tangential field-angle (deg.), and sagittal field-angle (deg.) associated with each of these four figures are summarized in Table 1. The same CCD model was used in all simulations. The array consisted of a 31-by-31 planar arrangement of square detectors measuring 0.01 mm on a side with a center-to-center spacing of 0.01 mm. Each point spread function was obtained by tracing 20,000 rays.

Table 1

	Spherical Aberration	Coma	Astigmatism	Defocus
Object plane distance	48.6	48.6	48.6	48.6
Image plane distance	52.0	52.0	50.4	51.0
Focal length	24.3	24.3	24.3	24.3
F-number	4.80	5.70	5.70	4.00
Tangential field-angle	0.00	2.87	5.91	0.00
Sagittal field-angle	0.00	2.87	0.00	0.00

On the top line, center diagram, of Figs. 2 through 5 are mesh plots of the four point spread functions considered. Each point spread function $p(x,y)$ is represented by a 31-by-31 matrix. The ideal extended object $o(x,y)$ used in this analysis was a 256-by-256 matrix which is displayed as an 8-bit gray-level diagram in the upper-left hand corner of each of these figures. Performing a two-dimensional convolution, see Eq. (35), of the extended object with the point spread function (in the absence of noise) yields a 286-by-286 degraded image $i(x,y)$ of which the central 256-by-256 portion of the degraded image is shown in the upper-right hand corner of each of these figures. The full 286-by-286 matrix is used, however, in subsequent image processing computations.

On line two of each of the four figures are mesh plots of the inverse point spread functions (31-by-31 matrices) $d_\beta(x,y)$ (left-diagram) and $w_\gamma(x,y)$ (right-diagram) obtained using the RAPID and Wiener filter algorithms, respectively. Performing a two-dimensional convolution of the point spread function $p(x,y)$ with each of the inverse point spread functions $d_\beta(x,y)$ and $w_\gamma(x,y)$ yields processed point spread functions (61-by-61 matrices). The central 31-by-31 portions of these processed point spread functions are shown as mesh plots on line three (left- and right-diagrams, respectively). For purposes of comparison, a 31-by-31 null-array with a single non-zero entry for the center pixel (Delta) is also displayed on line three, center diagram.

The values of the regularization parameters (β and γ) which gave rise to the best processed point spread functions, judged visually, for this analysis are: spherical aberration (0 and 0), coma ($3x10^{-2}$ and $1x10^{-4}$), astigmatism ($2x10^{-3}$ and $2x10^{-2}$), and defocus ($5x10^{-4}$ and $1x10^{-2}$). These same regularization parameter values were used in computing the object estimates $\hat{o}_\beta(x,y)$ and $\hat{o}_\gamma(x,y)$ using Eqs. (37) and (42), respectively. These object estimates (316-by-316 processed images) are displayed as gray-level diagrams on the bottom-line of each of the four figures. Only the central 256-by-256 portions of the processed images are shown.

The results presented in Figs. 2 through 5 were based on studies using synthesized point spread functions. We were fortunate to obtain real digitized degraded images of the planet Saturn taken with the wide-field planetary camera of the Hubble Space Telescope. The upper diagram in Fig. 6 shows a 400-by-250 degraded (unprocessed) image of the planet Saturn. The second line in Fig. 6 shows two 31-by-31 degraded images of different stars (called star #1 and star #2) also taken with the wide-field planetary camera. The RAPID algorithm was used to process the degraded image of Saturn using the two star images as point spread functions characterizing the degradation process. In particular, star #1 image was used as the input point spread function. Both star #1 and star #2 images were first processed using the inverse point spread function obtained using the star #1 image only. The regularization parameter, β, selected was the one which gave rise to processed star images of equal quality, based on a minimum entropy criterion. This same inverse point spread function was then used, via Eq.(35), to process the degraded Saturn image. The size of the reconstructed image was 430-by-280. The central 400-by-250 portion of this reconstruction is shown in the lower diagram of Fig. 6.

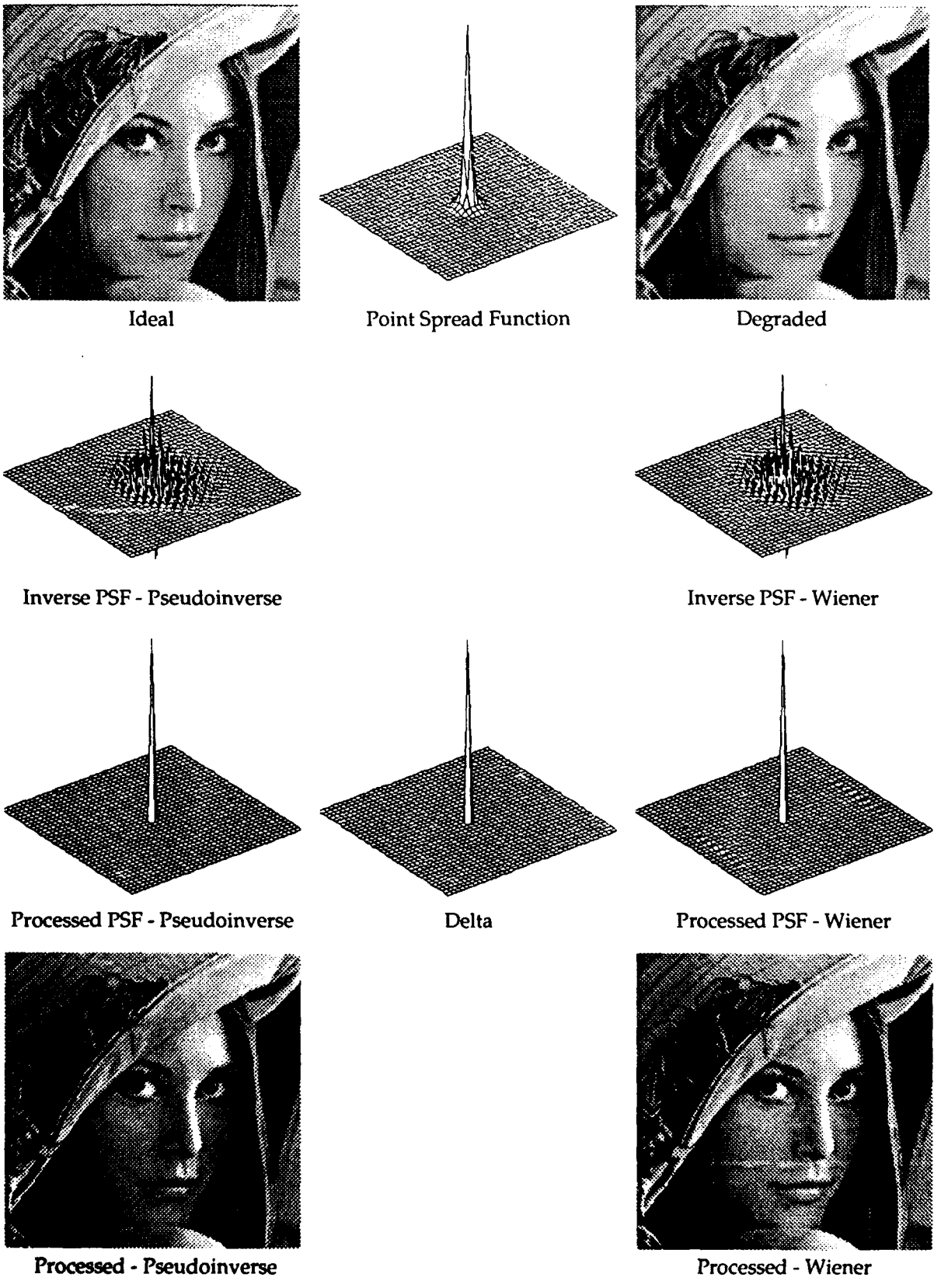
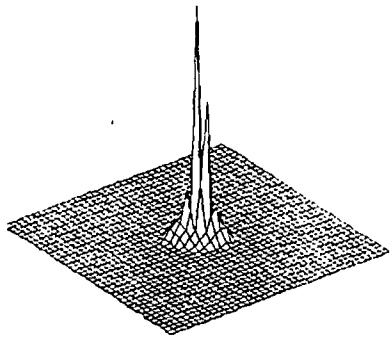


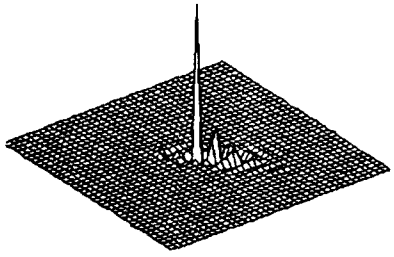
Figure 2. Point Spread Function Dominated by Spherical Aberration

Ideal

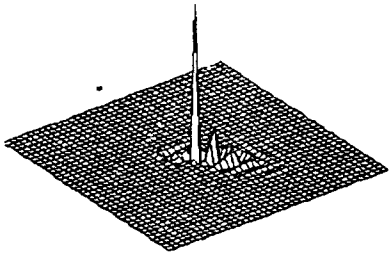


Point Spread Function

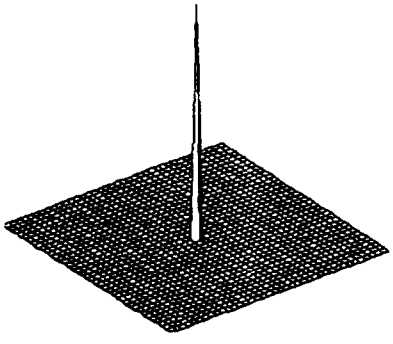
Degraded



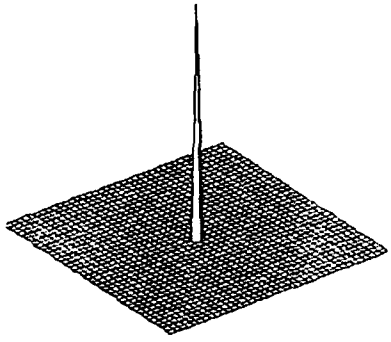
Inverse PSF - Pseudoinverse



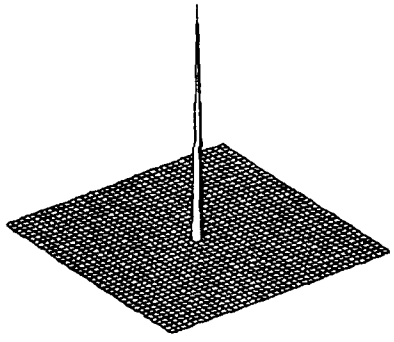
Inverse PSF - Wiener



Processed PSF - Pseudoinverse



Delta



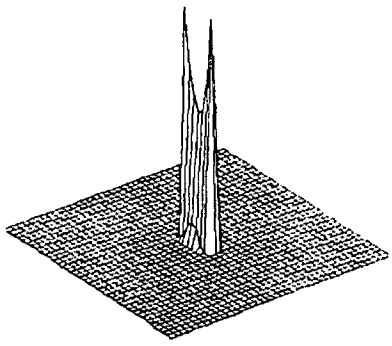
Processed PSF - Wiener

Processed - Pseudoinverse

Processed - Wiener

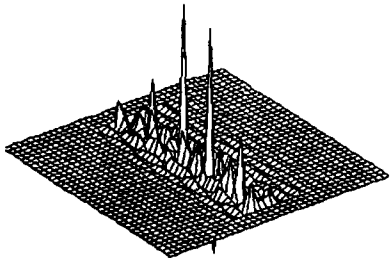
Figure 3. Point Spread Function Dominated by Coma

Ideal

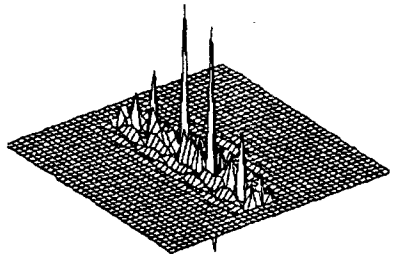


Point Spread Function

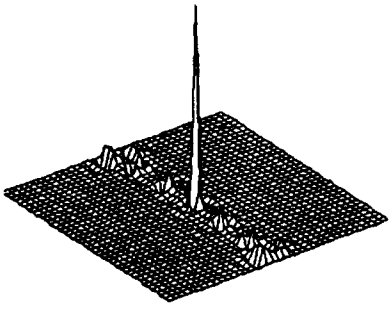
Degraded



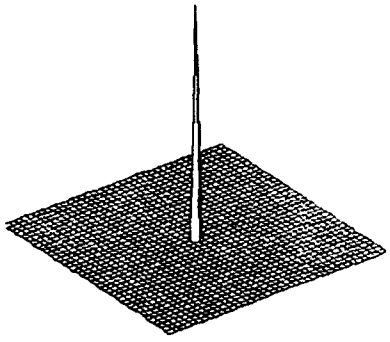
Inverse PSF - Pseudoinverse



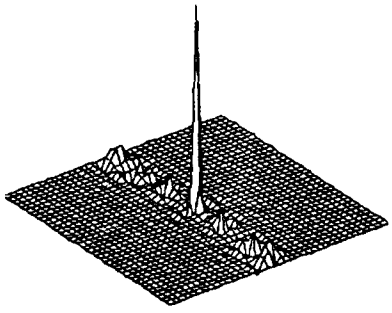
Inverse PSF - Wiener



Processed PSF - Pseudoinverse



Delta



Processed PSF - Wiener

Processed - Pseudoinverse

Processed - Wiener

Figure 4. Point Spread Function Dominated by Astigmatism

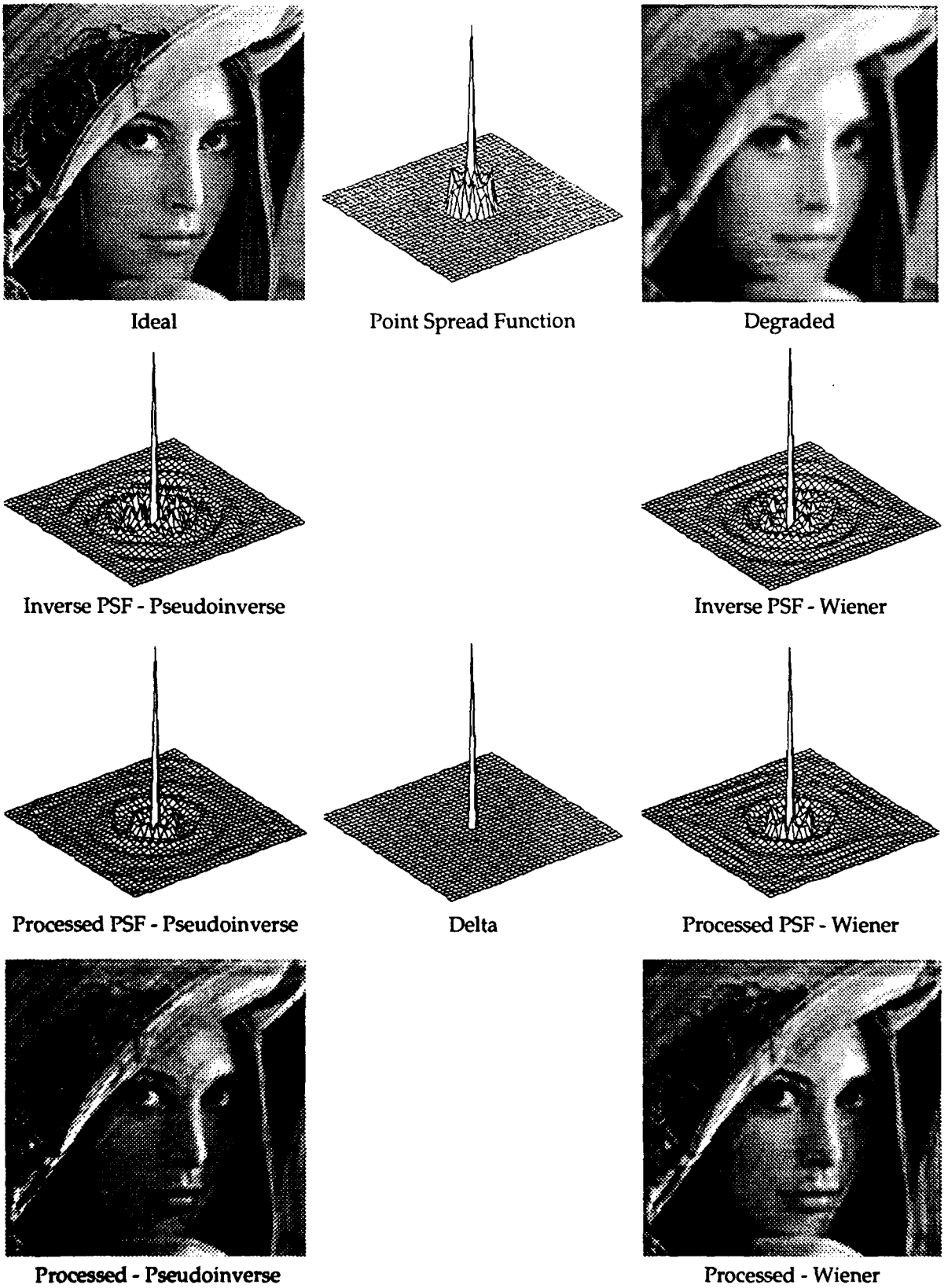
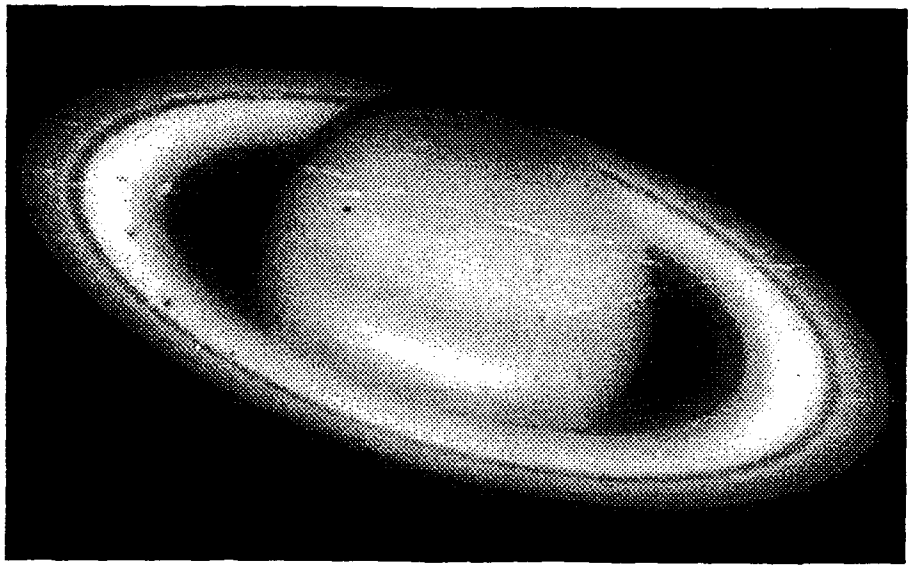
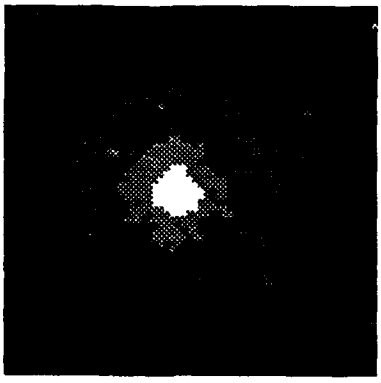


Figure 5. Point Spread Function Dominated by Defocus

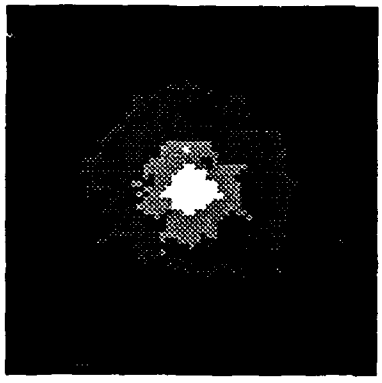


Unprocessed Saturn Image



Star #1 Image

Star #2 Image



Processed Saturn Image

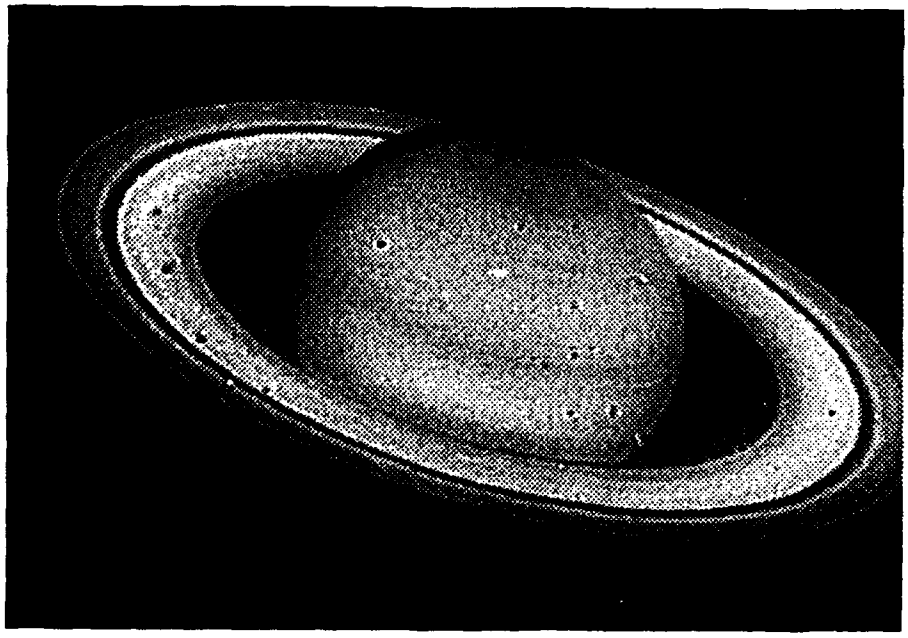


Figure 6. Hubble Space Telescope Imagery

9. CONCLUSIONS

A new algorithm, the Regularized Pseudoinverse Deconvolution (RAPID) algorithm, has been developed and has been shown to be equivalent in performance to Wiener filtering for shift-invariant point spread functions. The algorithm can be implemented either by direct convolution or indirectly using Fast Fourier Transform (FFT) techniques. The advantage of this approach is that, through the use of linear congruential scanning and the application of the circulant expansion of equation (31), regularized reconstruction methods can be applied to extended images degraded by shift-variant point spread functions.

10. ACKNOWLEDGEMENTS

This research was supported by the Strategic Defense Initiative Organization, Innovative Science and Technology Office, (SDIO/ISTO) through the Office of Naval Research and managed by Dr. Keith Bromley. Thanks are due to the National Aeronautics and Space Administration (NASA) for providing the digitized image of the planet Saturn obtained with the wide-field planetary camera of the Hubble Space Telescope. Special thanks go to Mr. Tom Phillips, Code 414, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, for developing the ray-trace program and supplying the point spread functions used in the computations.

11. REFERENCES

1. J. B. Abbiss and P. G. Earwicker, "Compact operator equations, regularization and super-resolution," *Proceedings, IMA Conference on Mathematics in Signal Processing*, University of Bath, Bath, United Kingdom, 17-19 September 1985. (Published by Clarendon Press, Oxford, U.K., 1987.)
2. D. J. Lindler and S. R. Heap, "Block iterative restoration of astronomical images with the NASA/Goddard massively parallel processor," *Proceedings, Second International Conference on Supercomputing*, (Lana P. Kartashev and Steven I. Kartashev, eds.), International Supercomputing Institute, Inc., 1987.
3. P. A. Jansson (Ed.), *Deconvolution with Applications in Spectroscopy*, Academic Press, New York, 1984.
4. C. K. Rushforth, "Signal restoration, functional analysis, and Fredholm integral equations of the first kind," in: H. Stark (Ed.), *Image Recovery: Theory and Application*, Academic Press, New York, 1987.
5. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, New Jersey, 1989.
6. M. A. Evgrafov, *Analytic Functions*, reprinted by Dover Publications, New York, 1978.
7. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington DC, 1977.
8. F. Smithies, *Integral Equations*, Cambridge Tract No. 49, Cambridge University Press, Cambridge, England, 1958.
9. B. R. Frieden, "Evaluation, design and extrapolation methods for optical signals, based on the use of the prolate functions," *Progress in Optics*, (E. Wolf, ed.), Vol. IX, pp. 313-407, North Holland, Amsterdam, 1971.
10. C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman Publishing Ltd., London, 1984.
11. K. Miller, "Least-squares methods for ill-posed problems with a prescribed bound," SIAM J. Math. Anal., Vol. 1, No. 1, pp. 52-74, 1970.
12. A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York and London, 1972.
13. G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, Vol. 14, No. 4, pp. 651-667, 1977.
14. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *Argonne National Laboratory Preprint MCS-P157-0690*, July 1990.
15. G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numer. Math.*, Vol. 14, pp. 403-420, 1970.
16. P. C. Hansen, "Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank," *SIAM J. Sci. Stat. Comput.*, Vol. 11, No. 3, pp. 503-518, May 1990.
17. P. J. Davis, *Circulant Matrices*, John Wiley & Sons, New York, 1979.
18. C.L.Matson, "Fourier spectrum extrapolation and enhancement using support-constraints," *Proc. SPIE, 1767* (1992).
19. J. M. Speiser and H. J. Whitehouse, "Multidimensional hybrid signal processing architectures", *Proceedings of the SPIE Advanced Institute on Transformations in Optical Signal Processing*, SPIE Vol. 373, 1981.
20. G. Ammar and P. Gader, "New decompositions of the inverse of a Toeplitz matrix," in: A.A. Kaaschock, J. H. van Schuppen and A. C. M. Ran (Eds.), *Signal Processing, Scattering and Operator Theory, and Numerical Methods*, Birkhauser, Boston, 1990.
21. R. C. Gonzalez and P. Wintz, *Digital Image Processing* (Second Edition), Addison-Wesley, Reading, Massachusetts, 1987.
22. G. H. Golub and C. F. Van Loan, *Matrix Computations* (Second Edition), The Johns Hopkins University Press, Baltimore, 1989.
23. T. Kailath, S.Y. Kung and M. Morf, "Displacement ranks of matrices and linear equations," *J. Math. Anal. and Appl.*, Vol. 68, pp. 395-407, 1979.
24. J. M. Speiser, H. J. Whitehouse and J. C. Allen, "Fast matrix-vector multiplication using displacement rank approximation via an SVD", *Proceedings Second International Workshop on SVD and Signal Processing*, University of Rhode Island, 25-27 June 1990.

APPENDIX A - TRADEOFFS BETWEEN ACCURACY AND SPEED

In this appendix we examine the RAPID reconstruction scheme in the continuous domain. An expression for the error displays a compromise between accuracy and speed. Let A denote the convolution operator on $L^2(\Re^2)$:

$$Af(x) = (a \star f)(x) = \int_{\Re^2} a(x-y) f(y)\, dm(y).$$

Let $\hat{}$ denote the two-dimensional Fourier transform and let $M_{\hat{a}}$ denote the multiplication operator: $M_{\hat{a}}\varphi = \hat{a}\phi$. As is well-known (see reference A1 below), the Fourier transform diagonalizes the convolution operator: $\hat{}^{-1} A \hat{} = M_{\hat{a}}$, with $\|A\| = \|\bar{a}\|_\infty$. In this formalism, the regularization operator satisfies:

$$\hat{}^{-1} A_\beta^+ \hat{} = M_{(|\hat{a}|^2 + \beta)^{-1}\bar{a}}.$$

The RAPID operator is obtained by determining the regularization pseudoinverse and then restricting it to a region near the origin. Introduce the associated convolution operator

$$A_b f(x) = \int_{\Re^2} (w_b a)(x-y) f(y)\, dm(y),$$

where w_b denotes the boxcar window: $w_b(x) = 1$ if $x \in [-b, b] \times [-b, b]$, zero otherwise. The regularization operator for A_β is

$$\hat{}^{-1} A_{b,\beta}^+ \hat{} = M_{w_b \star (|\hat{a}|^2 + \beta)^{-1}\bar{a}}.$$

Then the error between the regularization operator and the RAPID operator is

$$\left\| A_\beta^+ - A_{b,\beta}^+ \right\| = \left\| \frac{\bar{a}}{|\hat{a}|^2 + \beta} - w_b \star \frac{\bar{a}}{|\hat{a}|^2 + \beta} \right\|_\infty.$$

As b tends to infinity, the error tends to zero. However, a large value for b corresponds to an increased computational load. Thus, a trade off must be made between speed and accuracy. In addition, the error bound also indicates that a different window (such as a square Hamming) might improve the accuracy of the regularization scheme.

A1. N. Young, *An Introduction to Hilbert Space*, Cambridge University Press, 1988.

Paper presented at 3rd IMA Conference on Mathematics in Signal Processing,
University of Warwick, Warwick, UK, 15-17 December 1992.

FAST REGULARISED DECONVOLUTION IN OPTICS AND RADAR

John B. Abbiss

TITAN Spectron
1582 Parkway Loop
Suite B
Tustin, CA 92680-6505

Jeffery C. Allen, Richard P. Bocker, Harper J. Whitehouse

Naval Command, Control and Ocean Surveillance Center
RDT&E Division
San Diego, CA 92152-5000

ABSTRACT

In a paper presented at the first conference in this series, the problem was considered of reconstructing an object from image data degraded by a compact linear operator. Simulated synthetic aperture radar measurements were used to demonstrate that, using Tikhonov regularization techniques, resolution well beyond the conventional limit can be achieved. These image restoration methods have been further extended and are being investigated for application to space-based surveillance systems at optical wavelengths and also to laser radar imaging and detection problems. The computational burden associated with the reconstruction procedure is of critical importance, since the calculation must be performed rapidly enough to permit an effective response. For realistically-sized images, however, it can easily be shown that the operations count is prohibitively large, and some method of accelerating the deconvolution process must be found.

Approximations to the reconstruction matrix based on circulant expansions, which permit fast Fourier transform techniques to be exploited, represent a promising area for research. A new algorithm for fast restoration of extended images, the Regularized Pseudoinverse Deconvolution (RAPID) algorithm, suggested by the Toeplitz-like structure of the reconstruction matrix, is also proposed and discussed. Applications to both optical and radar systems are considered and illustrated. In the radar case, the squared modulus of the transmitted signal's ambiguity function plays the role of the optical point spread function.

1. INTRODUCTION

In the absence of any other degrading effects, the performance of an optical system is ultimately restricted by diffraction. The finite extent of the entrance pupil imposes a fundamental upper limit on the system's spatial frequency response. The image quality of most operational systems will not, however, approach this theoretical limit very closely. It is possible that the design or construction will be flawed, as in the case of the Hubble Space Telescope, through defective manufacture, assembly or quality assurance procedures. The methods of image restoration considered here were originally aimed at extending the performance of an optical system beyond the diffraction limit[1], but are in fact capable of compensating also for aberrations induced by the optical components. They are inherently robust and amenable to parallel implementations.

The assumption is made that the imaging system can be characterized by a point spread function which induces a possibly time- and shift-variant blurring of the image. It will be assumed that the point spread function is known or can be measured. In the case of the Hubble wide-field planetary camera, for example, there are two primary components to the point spread function; one due to diffraction by the aperture obstructions and the other due to spherical aberration. For a delay Doppler imaging radar, on the other hand, the point spread function is a bilinear functional of the transmitted signal.

The overall imaging system equation can be written as a Fredholm integral equation of the first kind. Solutions to ill-posed problems of this type are known to be numerically unstable[2, 3]. Additionally, it is anticipated that the image will be spatially sampled which will introduce discretization and associated noise processes. Thus, after scanning the image into a vector, the integral representing the continuous image can be rewritten as a matrix expression. In general, the presence of the sensor noise takes the measured image vector out of the span of the columns of the kernel matrix, which is typically highly ill-conditioned. Thus, when it is desired to estimate what the image would have been in the absence of aberrations or with less diffraction than the instrument can provide, techniques derived from regularization theory are required to restore stability to the reconstruction. By introducing a suitable error criterion (based on, *e.g.*, a vector norm), images can be constructed which are, in terms of the chosen criterion, closer to the undistorted geometrical image of the object than the detected image data.

To the extent permitted by the noise in the image, in-band effects can usually be removed by some form of pseudoinverse filter[4]. However, detector pixellation and the finite aperture of any system set fundamental resolution limits in the performance of such filters, and a method for achieving spectral extrapolation has to be devised. The spatial spectrum of the object is the Fourier transform of its amplitude, in the coherent case, or its intensity, in the incoherent case. If the object is known to be of finite extent, its Fourier transform is an analytic function, and the out-of-band part of the spectrum can in principle be fully recovered by analytic continuation[5] of the image spectrum after removal of any in-band distortion. The inverse Fourier transform of this extended spectrum would then yield a perfect image of the original object. Equivalently, one could attempt to solve directly the equation describing the imaging process. This, however, involves the inversion of an ill-conditioned matrix, and the restoration process is intrinsically unstable, even small amounts of noise rendering the results meaningless. These difficulties may be surmounted by applying the methods of regularization theory[6], developed to deal with ill-posed problems of this type; the solution is derived by means of a constrained least-squares procedure in which a regularization parameter plays an essential role. Stability in the restored image, which is computed via the regularized pseudoinverse of the imaging operator, is controlled by this parameter, whose optimal value depends on the signal-to-noise ratio in the data.

2. NATURE OF THE PROBLEM

We wish to estimate the object f from an image g, given that

$$g = Af + r \tag{1}$$

where A is the imaging operator and r represents the corrupting effect of additive noise. For clarity in the analysis, we consider the one-dimensional case with the operator A given in integral form by

$$(Af)(y) = \int_a^b A(x, y) f(x)\, dx, \qquad c \le y \le d. \tag{2}$$

In the absence of noise, Eq.(1) becomes a Fredholm equation of the first kind, in which the unknown function appears only under the integral sign. We can identify the sources of difficulty in solving this equation in the presence of noise or other perturbations (such as computer round-off error) by means of a singular function analysis[7].

We expand the kernel of the integral in terms of the singular functions $u_i(x)$ and $v_i(y)$, orthonormal systems in object and image spaces respectively, and the singular values σ_i:

$$A(x,y) = \sum_{i=1}^{\infty} \sigma_i \, u_i(x) \, v_i(y). \tag{3}$$

The object and image can be expanded in the forms

$$f(x) = \sum_{i=1}^{\infty} f_i \, u_i(x) \tag{4}$$

$$g(y) = \sum_{i=1}^{\infty} g_i \, v_i(y) \tag{5}$$

where the coefficients are related to $f(x)$ and $g(y)$ by the integral formulae

$$f_i = \int_a^b f(x) \, u_i(x) \, dx \tag{6}$$

and

$$g_i = \int_c^d g(y) \, v_i(y) \, dy. \tag{7}$$

In the noiseless case $(r = 0)$, we find from Eqs. (2) and (6)

$$(Af)(y) = \sum_{i=1}^{\infty} \sigma_i \, f_i \, v_i(y). \tag{8}$$

It follows, using Eqs.(5) and (8), that

$$g_i = \sigma_i \, f_i \tag{9}$$

and hence

$$f(x) = \sum_{i=1}^{\infty} \frac{g_i}{\sigma_i} \, u_i(x). \tag{10}$$

Thus the object function can in principle be perfectly reconstructed from the set $\{g_i\}$ of image coefficients.

Now consider the effects of noise. By expanding $r(y)$ in terms of the $v_i(y)$, we can derive the contribution of the noise to the new image coefficients:

$$g'_i = \sigma_i \, f_i + r_i. \tag{11}$$

The estimate $\hat{f}(x)$ is now

$$\hat{f}(x) = f(x) + \sum_{i=1}^{\infty} \frac{r_i}{\sigma_i} \, u_i(x). \tag{12}$$

Image formation is characteristically described by an integral transform of convolution type, *i.e.*, A is a convolution operator. Its singular-value spectrum typically decays asymptotically at an exponential rate[8]. Since the r_i will in general decrease less quickly, the sum in Eq.(12) will be divergent and no bound will exist for the 'distance' (in the sense of some appropriate metric) between the true object and the reconstruction. The effect of the noise on the reconstructed image is a manifestation of the fact that convolution is a strongly smoothing process - closely similar images can correspond to widely differing objects. Thus image restoration is an ill-posed problem, small perturbations in the data causing large changes in the solution represented by Eq.(12).

3. THE REGULARIZED SOLUTION

Methods of regularization theory[6, 9] can be exploited to convert the problem to a related well-posed one, *i.e.*, one for which the solution exists, is unique and depends continuously on the data. Since we shall later be concerned with computed reconstructions, we henceforth consider the problem in its finite discretized form; thus the imaging operator A becomes a matrix. Although strictly we should introduce new symbols, for convenience we continue to use f, g, and A to denote the discrete forms of object, image, and imaging operator. Generally $f \in C^n$, $g \in C^m$ and $A \in C^{m \times n}$.

To regularize the problem, we shall modify A. We impose constraints[10] on possible solutions f' by requiring that

$$\| Af' - g \|^2 \leq \varepsilon^2 \tag{13}$$

where ε is some suitable measure of the noise in the image, and that

$$\| f' \|^2 \leq \xi^2 \tag{14}$$

where ξ is some suitable measure of the permitted 'signal strength' of the solution. ($\| \bullet \|$ denotes norm in the Hilbert spaces associated with object and image.) We combine these constraints and minimize

$$\| Af' - g \|^2 + \beta \| f' \|^2$$

where the regularization parameter β is given by

$$\beta = \varepsilon^2 / \xi^2 \ . \tag{15}$$

The minimum-norm solution to this constrained least-squares problem is given by

$$f_\beta = A_\beta^+ g \tag{16}$$

where

$$A_\beta^+ = (A^H A + \beta I)^{-1} A^H . \tag{17}$$

We note the relationship of A_β^+ (which we shall call the regularized pseudoinverse) to A^+, the Moore-Penrose pseudoinverse[11]

$$A^+ = \lim_{\beta \to 0} A_\beta^+ . \tag{18}$$

The inverse of $(A^H A + \beta I)$ always exists, since $A^H A$ is non-negative definite and the regularization parameter β is positive. A value of β should be chosen which balances data fidelity against smoothness in the reconstruction. Methods are also available for determining β from the image data themselves[12, 13].

4. SINGULAR VALUE AND FOURIER DECOMPOSITIONS

It is often convenient to compute the regularized pseudoinverse via the singular value decomposition (SVD) of A:

$$A = U\Sigma V^H \tag{19}$$

where[14]

$$U^H U = I_m, \qquad V^H V = VV^H = I_n \tag{20}$$

and

$$\Sigma = diag\,(\sigma_1, \sigma_2, \ldots, \sigma_n), \qquad \sigma_i \geq 0. \tag{21}$$

The singular values $\{\sigma_i\}$ are assumed to have been arranged in descending order of magnitude

$$\sigma_1 > \sigma_2 > \sigma_3 > \ldots > \sigma_n. \tag{22}$$

Then we find

$$f_\beta = V\ \Sigma_\beta^+\ U^H\ g \tag{23}$$

where

$$\Sigma_\beta^+ = diag\left(\ldots, \frac{\sigma_i}{\sigma_i^2 + \beta}, \ldots\right). \tag{24}$$

This representation is useful when the behavior of the reconstruction as a function of the regularization parameter is being studied. Regularization can equivalently be achieved by setting β to zero, and simply truncating the singular value series at a point which is dependent on the noise level[15].

An *ab initio* computation of f_β via Eq.(23) requires the SVD of A followed by two matrix-vector multiplications. If the regularized pseudoinverse can be precomputed, only one matrix-vector product is needed to generate the reconstruction. For images of more than modest sizes, however, the computation rapidly becomes burdensome. If, for example, f and g are 100-by-100, A is a 10^4-by-10^4 matrix, and the matrix-vector product requires 10^8 multiplications. There will also be considerable storage demands. Thus, if major computational resources are not available, some means of simplifying the calculation will be needed in many cases of practical interest.

If the matrix A were square circulant of order n ($A = [a_{j-i+1}]$, subscript mod n), we could dramatically reduce the computational burden by exploiting the fact that the Fourier transform diagonalizes a circulant[16]. If F denotes the Fourier matrix:

$$F^H = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & w & w^2 & \ldots & w^{n-1} \\ 1 & w^2 & w^4 & \ldots & w^{2(n-1)} \\ & & & \ldots & \\ 1 & w^{n-1} & w^{2(n-1)} & \ldots & w^{(n-1)^2} \end{bmatrix}, \qquad w = exp\,(i2\pi/n) \tag{25}$$

then

$$A = F^H \Lambda F \tag{26}$$

where

$$\Lambda = diag\,(\lambda_1, \lambda_2, \ldots, \lambda_n). \tag{27}$$

It follows from Eq.(16) that

$$f_\beta = F^H \Lambda_\beta^+ F\, g \tag{28}$$

where

$$\Lambda_\beta^+ = diag\left(\ldots, \frac{\lambda_i^*}{|\lambda_i|^2 + \beta}, \ldots\right) \tag{29}$$

and λ_i^* is the complex conjugate of λ_i. Note that the singular values $\{\sigma_i\}$ and the eigenvalues $\{\lambda_i\}$ are related by

$$\sigma_i = |\lambda_i|. \tag{30}$$

The operational counts for the SVD and the FFT are $O(n^3)$ and $O(n \log n)$, respectively.

5. STRUCTURE OF THE IMAGING MATRIX

Under some circumstances the imaging matrix can be readily modified to circulant form. For a shift-invariant one-dimensional system, the image is a convolution of the point spread function and the object. If the sampling intervals in image and reconstruction spaces are equal, the matrix A is then Toeplitz ($A = [a_{j-i}]$). If, in addition, the image and reconstruction vectors are of equal length, A can be padded to circulant form. In two dimensions, f and g are matrices and must be mapped into vectors. The manner in which this is performed will determine the structure of A. For column-wise mapping, for instance, again with equal sampling in image and reconstruction spaces, A becomes block-Toeplitz with Toeplitz blocks. If the image and reconstruction matrices have the same number of elements, A can be padded to become block-circulant with circulant blocks, and the problem is again amenable to Fourier transform methods. In both one and two dimensions, it should be noted that the penalty associated with the expansion of A to circulant or block-circulant form is a relaxation of the support constraint on the reconstruction, which renders the calculation, and in particular the degree of resolution enhancement achieved, much more sensitive to noise[17]. An alternative construction in the two-dimensional case is to zero-pad f and g into larger matrices and then to use a linear congruential scan to map the padded matrices into vectors. A then becomes a circulant matrix since the linear congruential scan is an isomorphism between 2D convolution and 1D convolution[18].

For less structured imaging matrices (*e.g.*, if the system is weakly space-variant) it may be asked whether accelerated computation of the matrix-vector product is still possible. In this context, recent work on circulant approximations to matrices of quite general form[19] appears highly relevant, and includes the following result. For any matrix, A_β^+ say, we can write

$$A_\beta^+ = C_0 + \sum_{m=1}^{\alpha} L(x_m)\, C^T(y_m) \tag{31}$$

where C_o is a circulant matrix with the same last row as A_β^+, $L(x_m)$ is a lower triangular Toeplitz matrix with x_m as its first column, and $C(y_m)$ is a circulant matrix whose last row is y_m. The $\{x_m\}$ and $\{y_m\}$ may be obtained from the truncated SVD of the cyclic displacement of A_β^+:

$$A_\beta^+ - EA_\beta^+E^T = \sum_{m=1}^{\alpha} x_m y_m^T \tag{32}$$

where E is the cyclic downshift matrix

$$E = \begin{bmatrix} 0\,0\,0\,\ldots\,0\,1 \\ 1\,0\,0\,\ldots\,0\,0 \\ 0\,1\,0\,\ldots\,0\,0 \\ \qquad\ldots \\ 0\,0\,0\,\ldots\,1\,0 \end{bmatrix}. \tag{33}$$

For imaging matrices with strongly Toeplitz features, α should be a small number.

6. THE REGULARIZED PSEUDOINVERSE DECONVOLUTION ALGORITHM

Consider a two-dimensional optical imaging system whose point spread function is both time- and space-invariant. In the presence of additive noise, the imaging equation connecting the input (extended object), the output (degraded image), and the point spread function (impulse response) is given by the following two-dimensional convolutional integral equation

$$i(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} o(x',y')\,p(x-x',y-y')\,dx'dy' + n(x,y). \tag{34}$$

In shorthand notation

$$i = p \bigstar\bigstar o + n \tag{35}$$

where $o(x,y)$ represents the extended object, $i(x,y)$ the degraded image, $p(x,y)$ the point spread function, and $n(x,y)$ the noise. It is assumed that the noise is independent of position in the image.

The discrete version of Eq.(34) can, of course, be cast[20] into the following vector-matrix form

$$g = Af + r. \tag{36}$$

When, in particular, f, g, and r represent the one-dimensional column vectors formed by stacking the rows or columns of the discretized versions of the input $o(x,y)$, output $i(x,y)$, and the noise $n(x,y)$, respectively, the two-dimensional imaging matrix A is block-Toeplitz with Toeplitz blocks. It can be shown, using known results [21] that the regularized pseudoinverse A_β^+ is block-persymmetric with persymmetric blocks. The inverse of a Toeplitz matrix is persymmetric, and persymmetric matrices obtained by inverting Toeplitz matrices have much more Toeplitz-like structure than general persymmetric matrices. In particular, their displacement rank is the same as that of the parent-Toeplitz matrix[22, 23]. Displacement rank, it should be noted, is a quantitative measure of the closeness of a given matrix to being Toeplitz. In one dimension, the displacement rank of a Toeplitz matrix is ≤ 2, and the displacement rank of A_β^+ is ≤ 4.

In the early stages of this investigation, it was noted that for a variety of point spread functions being studied the corresponding computer-generated regularized pseudoinverse matrices appeared to have banded block-Toeplitz structure. The implications of this observation were considered. In particular, if a space-invariant point spread function used in a two-dimensional linear convolutional imaging equation leads to a block-Toeplitz imaging matrix with Toeplitz blocks, then the converse must also be true. That is, given a block-Toeplitz imaging matrix containing Toeplitz blocks, then the corresponding space-invariant point spread function which gave rise to this imaging matrix could be easily ascertained. This implies that from the regularized pseudoinverse an inverse point spread function $d_\beta(x, y)$ could be constructed which could be used to process the image $i(x,y)$ and form an estimate $\hat{o}_\beta(x, y)$ of the original object. This two-dimensional linear convolution technique is summarized by the equation

$$\hat{o}_\beta = d_\beta \star\star i. \tag{37}$$

The technique was tested on a digital computer with encouraging results. Further experimental investigations indicate that results obtained with this Regularized Pseudoinverse Deconvolution (RAPID) algorithm are comparable in quality to those obtained using parametric Wiener filtering. Results using degraded images processed with both the RAPID algorithm and parametric Wiener filtering are presented in section 8.

7. DECONVOLUTION VIA WIENER FILTERING

Wiener filtering[20] is a well-known technique for processing images degraded and corrupted by noise as described by Eq.(34). From a knowledge of the point spread function $p(x,y)$ characterizing the imaging system, it is possible to compute the corresponding transfer function $\tilde{p}(f_x, f_y)$ using the two-dimensional Fourier transform

$$\tilde{p}(f_x, f_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)\, exp\left[-i2\pi(xf_x + yf_y)\right] dx dy. \tag{38}$$

From the transfer function and knowledge of the power spectra of object $S_o(f_x, f_y)$ and noise $S_n(f_x, f_y)$, the parametric Wiener filter can be constructed, namely

$$\tilde{w}_\gamma(f_x, f_y) = \frac{\left|\tilde{p}(f_x, f_y)\right|^2}{\tilde{p}(f_x, f_y)\left[\left|\tilde{p}(f_x, f_y)\right|^2 + \gamma\left|\dfrac{S_n(f_x, f_y)}{S_o(f_x, f_y)}\right|\right]}. \tag{39}$$

When $\gamma = 1$, Eq.(39) reduces simply to the Wiener filter. If γ is variable we refer to this as the parametric Wiener filter. In the absence of noise, either form of the Wiener filter reduces to the ideal inverse filter. When the power spectra are not known, which is often the case in practice, Eq.(39) can be approximated by

$$\tilde{w}_\gamma(f_x, f_y) \approx \frac{\left|\tilde{p}(f_x, f_y)\right|^2}{\tilde{p}(f_x, f_y)\left[\left|\tilde{p}(f_x, f_y)\right|^2 + \gamma\right]}. \tag{40}$$

From the Wiener filter, using either Eqs.(39) or (40), an inverse point spread function $w_\gamma(x, y)$ can be constructed using the two-dimensional inverse Fourier transform. That is,

$$w_\gamma(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{w}_\gamma(f_x, f_y)\, exp\left[+i2\pi\left(xf_x + yf_y\right)\right] df_x df_y. \tag{41}$$

With the inverse point spread function given by Eq.(41), an estimate $\hat{o}_\gamma(x, y)$ of the object can be computed using the two-dimensional linear convolutional equation

$$\hat{o}_\gamma = w_\gamma \star\star i \tag{42}$$

where, again, $i(x,y)$ is the degraded image.

8. APPLICATIONS TO THE OPTICAL CASE

Preliminary results obtained for optical systems using the RAPID algorithm have been previously presented[24] and are summarized here. For purposes of comparison, some results obtained using the parametric Wiener filter algorithm are also included. The optical system considered for these studies was a simple spherical converging lens. The object and image planes are parallel and orthogonal to the optical axis of the lens. An arbitrary point source located in the object plane gives rise to an intensity distribution (the point spread function) in the image plane. A charge-coupled device (CCD), for example, can be used to measure the point spread function. A ray-trace program was used in the synthesis of two different point spread functions dominated by astigmatism and defocus. The object plane distance (mm), image plane distance (mm), focal length (mm), F-number, tangential field-angle (deg.), and sagittal field-angle (deg.) associated with these point spread functions are summarized in Table 1. The same CCD model was used in both simulations. The array consisted of a 31-by-31 planar arrangement of square detectors measuring 0.01 mm on a side with a center-to-center spacing of 0.01 mm. Each point spread function was obtained by tracing 20,000 rays.

Table 1

	Astigmatism	Defocus
Object plane distance	48.6	48.6
Image plane distance	50.4	51.0
Focal length	24.3	24.3
F-number	5.70	4.00
Tangential field-angle	5.91	0.00
Sagittal field-angle	0.00	0.00

On the top line, center diagram, of Figs.1 and 2 are mesh plots of the two of the point spread functions considered. Each point spread function $p(x,y)$ is represented by a 31-by-31 matrix. The ideal extended object $o(x,y)$ used in this analysis was a 256-by-256 matrix which is displayed as an 8-bit gray-level diagram in the upper-left hand corner of each of these figures. Performing a two-dimensional convolution, see Eq.(35), of the extended object with the point spread function (in the absence of noise) yields a 286-by-286 degraded image $i(x,y)$ of which the central 256-by-256 portion of the degraded image is shown in the upper-right hand corner of each of these figures. The full 286-by-286 matrix is used, however, in subsequent image processing computations.

On line two of each of these figures are mesh plots of the inverse point spread functions (31-by-31 matrices) $d_\beta(x,y)$ (left-diagram) and $w_\gamma(x,y)$ (right-diagram) obtained using the RAPID and Wiener filter algorithms, respectively. Performing a two-dimensional convolution of the point spread function $p(x,y)$ with each of the inverse point spread functions $d_\beta(x,y)$ and $w_\gamma(x,y)$ yields processed point spread functions (61-by-61 matrices). The central 31-by-31 portions of these processed point spread functions are shown as mesh plots on line three (left- and right-diagrams, respectively). For purposes of comparison, a 31-by-31 null-array with a single non-zero entry for the center pixel (Delta) is also displayed on line three, center diagram.

The values of the regularization parameters, β and γ, which gave rise to the best processed point spread functions, judged visually, for this analysis are: astigmatism, $2x10^{-3}$ and $2x10^{-2}$, and defocus, $5x10^{-4}$ and $1x10^{-2}$. These same regularization parameter values were used in computing the object estimates $\hat{o}_\beta(x,y)$ and $\hat{o}_\gamma(x,y)$ using Eqs. (37) and (42), respectively. These object estimates (316-by-316 processed images) are displayed as gray-level diagrams on the bottom-line of each of these figures. Only the central 256-by-256 portions of the processed images are shown.

The results presented in Figs. 1 and 2 were based on studies using synthesized point spread functions. We were fortunate to obtain real digitized degraded images of the planet Saturn taken with the wide-field planetary camera of the Hubble Space Telescope. The upper diagram in Fig. 3 shows a 400-by-250 degraded (unprocessed) image of the planet Saturn. Here the point spread function is weakly space-variant. Approximate allowance can be made for its variation across the image of Saturn by using the images of two stars in the same part of the field-of-view, obtained with the telescope on another occasion. The second line in Fig. 3 shows the 31-by-31 degraded images of these two stars (called star #1 and star #2). The RAPID algorithm was used to process the degraded image of Saturn using the star images as point spread functions characterizing the degradation process. In particular, star #1 image was used as the input point spread function. Both star #1 and star #2 images were first processed using the inverse point spread function obtained using the star #1 image only. The regularization parameter, β, selected was the one which gave rise to processed star images of equal quality, based on a minimum entropy criterion. The entropy of the normalized image is $-\sum\sum p_{ij} log p_{ij}$ where p_{ij} is the content of the i,j *th* pixel. This same inverse point spread function was then used, via Eq.(37), to process the degraded Saturn image. The size of the reconstructed image was 430-by-280. The central 400-by-250 portion of this reconstruction is shown in the lower diagram of Fig.3.

9. RADAR IMAGE RECONSTRUCTION

Microwave radar signal processing is a well-established field[25] in which the role of the matched filter is pre-eminent. More recently, microwave imaging radars have become important for terrain mapping not only of the earth[26] but also of other planetary objects[27]. Most terrain-mapping radars are of the synthetic aperture type; the ground is assumed to be stationary and the radar is assumed to move relative to it with a known motion, so that the radar return can be processed simultaneously in range and Doppler. After pulse compression, radar range (slant range) can be converted to range along the ground using the assumed orientation of the radar with respect to the ground. After

Doppler processing, cross-range can be computed from the processed signals. In a simple side-looking synthetic aperture radar, such as is used in commercial aircraft ground mapping, range and Doppler processing may be performed independently and the point spread function of the imaging radar treated as separable. Under these conditions, image formation can be viewed as a Kronecker product of two one-dimensional processes. In a paper presented at the first conference in this series[28], an image restoration technique for a synthetic aperture radar system, based on the regularized pseudoinverse of the imaging operator, was considered.

In this paper, we discuss an inverse synthetic aperture laser radar example, for which the target will be modeled as a doubly-spread (in range and Doppler) extended object. We will initially assume that a monostatic laser radar system is looking into space at a distribution of reflecting spheres of various sizes at different ranges and associated with different Doppler frequencies. It is reasonable to assume also that the target of the laser radar is rough relative to the illuminating radiation wavelength. This assumption is approximately true even for microwave radars operating in a planetary-mapping mode[27]. Following Van Trees' standard statistical radar/sonar model[29], the return is assumed to be a Gaussian random variable. The detection statistic used is the log-likelihood function and is equal to the magnitude squared output of a bandpass-matched filter or correlator.

For those unfamiliar with the mathematical formalism describing the properties and performance of a radar system, we introduce some terminology. The traditional approach to modeling the radar image is via the asymmetric ambiguity function[29]:

$$\chi(\nu,\tau) = \int_{-\infty}^{\infty} z^*(t)\, z(t+\tau)\, exp(+i2\pi\nu t)\, dt \tag{43}$$

where $z(t) = s(t) + i\hat{s}(t)$ and $\hat{s}(t)$ denotes the Hilbert transform of $s(t)$ and τ and ν represent delay and Doppler shift. This narrowband ambiguity function may be viewed as the response of either a matched filter receiver to a point scatterer as a function of the target delay and Doppler shift, or a bank of matched filters to a point scatterer at one particular delay and Doppler shift. The symmetric form of this ambiguity function is

$$A(\nu,\tau) = exp(+i\pi\nu\tau)\,\chi(\nu,\tau) = \int_{-\infty}^{\infty} z^*\left(t-\frac{\tau}{2}\right) z\left(t+\frac{\tau}{2}\right) exp(+i2\pi\nu t)\, dt. \tag{44}$$

Feig and Grünbaum[30] have made use of the symmetric ambiguity function to demonstrate the connection between radar detection and tomography. Bernfeld[31] first recognized the analogy between the ambiguity function (more precisely, its squared modulus) and certain mathematical entities appearing in the equations governing x-ray tomographic data. Snyder *et al.*[32] noted the analogy with time-of-flight positron emission tomography. [See also ref. 33.] Whitehouse and Boashash[34] established a related analogy between the radar image and that obtained by time-of-flight positron emission tomography for the bistatic case, where the phase of the transmitted signal is not needed.

The Wigner-Ville distribution (WVD) of a real signal $s(t)$, associated with the complex analytic signal $z(t)$, is a time-frequency distribution defined as[35]

$$W(t,f) = \int_{-\infty}^{\infty} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) exp(-i2\pi f\tau)\, d\tau. \tag{45}$$

Note that the WVD is the double Fourier transform of the symmetric ambiguity function. The WVD is always real, whereas the ambiguity function is in general complex. For a given signal, the WVD is related to the squared modulus of the ambiguity function by a double convolution[35]:

$$W \star\star W = |A|^2. \tag{46}$$

As stated above, the radar return is assumed to be a Gaussian random variable, and the detection statistic used is the log-likelihood function. Under this stochastic treatment of the radar reflection, the determination of the target's scattering function σ becomes a problem in deconvolution:

$$E(|\chi_{sr}|^2) = |\chi_{ss}|^2 \star\star \sigma \tag{47}$$

where E is the expectation operator, s denotes the transmitted signal, r denotes the received signal and χ_{sr} is the cross-ambiguity function between transmitted and received signals:

$$\chi_{sr}(\nu, \tau) = \int_{-\infty}^{\infty} s^*(t)\, r(t+\tau) \exp(+i2\pi\nu t)\, dt. \tag{48}$$

Thus we can see from Eq.(47) that the known function $|\chi_{ss}|^2$ plays the role of the point spread function in the radar imaging case. The radar deconvolution problem can also be stated in the closely-related form[36]:

$$E(W_r) = W_s \star\star \sigma \tag{49}$$

where W_s and W_r are the WVDs of the transmitted and received signals respectively. A consistent estimator for this equation is discussed in Appendix A.

We observe that the total energy, $\mathcal{E}$, in the ambiguity function is

$$\mathcal{E} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |A(\nu, \tau)|^2 d\nu d\tau \tag{50}$$

and that $|A|^2$ possesses the self-transform property[25]:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |A(\nu, \tau)|^2 \exp[i2\pi(\tau f - \nu t)]\, d\nu d\tau = |A(t, f)|^2. \tag{51}$$

We also note an important distinction between the optical and radar cases. The resolution of a corrected optical system of modest F-number (ratio of focal length to objective diameter) can be improved simply by increasing the aperture. Suppose an iris is illuminated by a point source whose radiant intensity can be adjusted to maintain the power passing through the aperture at a constant level. As the diameter of the iris increases, the central peak of the point spread function [37] becomes both higher and narrower; the height increases in proportion to the area of the aperture while the width decreases in proportion to its diameter. Note that the contribution of the volume of the central peak relative to the integrated sidelobes remains constant. For the radar system, on the other hand, it can be shown that, for a constant-energy signal of increasing time-bandwidth product, the width of the central peak in the ambiguity function decreases, while its magnitude remains constant and the integrated volume under the sidelobes increases.

Thus, radar systems become increasingly aberrated as the time-bandwidth product increases, and we are obliged to resort to post-processing of the data to maintain feature contrast as we increase the resolution for distributed objects. This situation brings to mind the Hubble wide-field planetary camera, whose design resolution was achieved, but whose performance was degraded by the enlargement of the point spread function sidelobes through severe spherical aberration.

10. APPLICATION TO THE RADAR CASE

As an illustration of the application of the ideas discussed here to the radar case, we consider the ambiguity function associated with a laser radar system[38] radiating pulses with Gaussian envelopes. The ambiguity function is then a two-dimensional Gaussian. Since the WVD of the signal is also a two-dimensional Gaussian, the computation involved in the image restoration procedure has exactly the same form for both Eqs.(47) and (49). The computer software used in generating the simulation results in Figs. 1 and 2 was applied to the case of a circularly symmetric Gaussian point spread function. The results of this simulation are summarized in Fig.4.

11. ACKNOWLEDGEMENTS

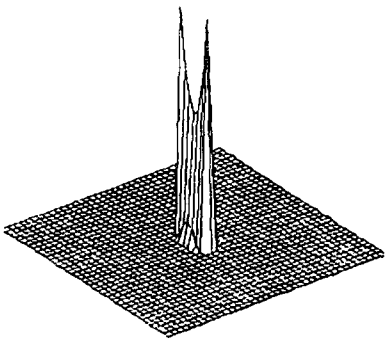
This research was supported by the Strategic Defense Initiative Organization's Innovative Science and Technology Office through the Office of Naval Research and managed by Dr. Keith Bromley. Thanks are due to the National Aeronautics and Space Administration for providing the digitized image of the planet Saturn obtained with the wide-field planetary camera of the Hubble Space Telescope. Special thanks go to Mr. Tom Phillips, Code 414, Naval Command, Control and Ocean Surveillance Center, RDT&E Division, for developing the ray-trace program and supplying the point spread functions used in the computations.

12. REFERENCES

1. J. B. Abbiss, C. De Mol and H. S. Dhadwal, "Regularized iterative and noniterative procedures for object restoration from experimental data", *Optica Acta*, Vol. 30, No.1, pp. 107-124, 1983.
2. P. A. Jansson (Ed.), *Deconvolution with Applications in Spectroscopy*, Academic Press, New York, 1984.
3. C. K. Rushforth, "Signal restoration, functional analysis, and Fredholm integral equations of the first kind", in: H. Stark (Ed.), *Image Recovery: Theory and Application*, Academic Press, New York, 1987.
4. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, New Jersey, 1989.
5. M. A. Evgrafov, *Analytic Functions*, reprinted by Dover Publications, New York, 1978.
6. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington DC, 1977.
7. F. Smithies, *Integral Equations*, Cambridge Tract No. 49, Cambridge University Press, Cambridge, England, 1958.
8. B. R. Frieden, "Evaluation, design and extrapolation methods for optical signals, based on the use of the prolate functions", *Progress in Optics*, (E. Wolf, ed.), Vol. IX, pp. 313-407, North Holland, Amsterdam, 1971.
9. C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman Publishing Ltd., London, 1984.
10. K. Miller, "Least-squares methods for ill-posed problems with a prescribed bound", *SIAM J. Math. Anal.*, Vol. 1, No. 1, pp. 52-74, 1970.
11. A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York and London, 1972.
12. G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy", *SIAM J. Numer. Anal.*, Vol. 14, No. 4, pp. 651-667, 1977.
13. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve", *Argonne National Laboratory Preprint MCS-P157-0690*, July 1990.
14. G. H. Golub and C. Reinsch, "Singular value decomposition and least-squares solutions", *Numer. Math.*, Vol. 14, pp. 403-420, 1970.

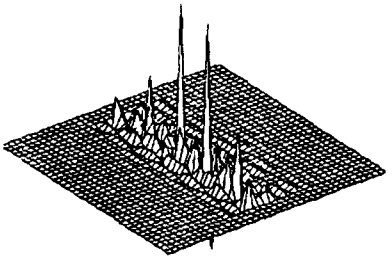
15. P. C. Hansen, "Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank", *SIAM J. Sci. Stat. Comput.*, Vol. 11, No. 3, pp. 503-518, May 1990.
16. P. J. Davis, *Circulant Matrices*, John Wiley & Sons, New York, 1979.
17. C. L. Matson, "Fourier spectrum extrapolation and enhancement using support-constraints", in *Inverse Problems in Scattering and Imaging*, Michael A. Fiddy, Editor, Proc. SPIE Vol. 1767, pp. 419-430, 1992.
18. J. M. Speiser and H. J. Whitehouse, "Multidimensional hybrid signal processing architectures", *Proceedings of the SPIE Advanced Institute on Transformations in Optical Signal Processing*, SPIE Vol. 373, 1981.
19. G. Ammar and P. Gader, "New decompositions of the inverse of a Toeplitz matrix", in: A.A. Kaaschock, J. H. van Schuppen and A. C. M. Ran (Eds.), *Signal Processing, Scattering and Operator Theory, and Numerical Methods*, Birkhauser, Boston, 1990.
20. R. C. Gonzalez and P. Wintz, *Digital Image Processing* (Second Edition), Addison-Wesley, Reading, Massachusetts, 1987.
21. G. H. Golub and C. F. Van Loan, *Matrix Computations* (Second Edition), The Johns Hopkins University Press, Baltimore, 1989.
22. T. Kailath, S.Y. Kung and M. Morf, "Displacement ranks of matrices and linear equations", *J. Math. Anal. and Appl.*, Vol. 68, pp. 395-407, 1979.
23. J. M. Speiser, H. J. Whitehouse and J. C. Allen, "Fast matrix-vector multiplication using displacement rank approximation via an SVD", *Proceedings Second International Workshop on SVD and Signal Processing*, University of Rhode Island, 25-27 June 1990.
24. J. B. Abbiss, J. C. Allen, R. P. Bocker and H. J. Whitehouse, "Fast image reconstruction based on the regularized pseudoinverse of the imaging operator", in *Inverse Problems in Scattering and Imaging*, Michael A. Fiddy, Editor, Proc. SPIE Vol. 1767, pp. 93-111, 1992.
25. M. I. Skolnik, *Radar Handbook*, McGraw-Hill Book Company, New York, 1970.
26. J. J. Kovaly, *Synthetic Aperture Radar*, Artech House, Dedham, MA, 1976.
27. J. V. Evans and T. Hagfors, *Radar Astronomy*, McGraw-Hill Book Company, New York, 1968.
28. J. B. Abbiss and P. G. Earwicker, "Compact operator equations, regularization and super-resolution", in *Mathematics in Signal Processing* (ed. J. G. McWhirter), Clarendon Press, Oxford, UK, 1987.
29. H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part III: Radar-Sonar Signal Processing and Gaussian Signals in Noise, John Wiley and Sons, New York, 1971.
30. E. Feig and F. A. Grünbaum, "Tomographic methods in range-Doppler radar", *Inverse Problems*, Vol. 2, pp. 185-195, 1986.
31. M. Bernfeld, "Chirp Doppler radar", *Proc. IEEE Letters*, Vol. 72, No. 4, pp. 540-541, 1984.
32. D. L. Snyder, L. J. Thomas and M. M. TerPogossian, "A mathematical model for positron emission tomography systems having time-of-flight measurements", *IEEE Trans. Nuclear Science*, Vol. NS-28, pp. 3575-3583, 1981.
33. D. L. Snyder and H. J. Whitehouse, "Delay-Doppler radar imaging using chirp-rate modulation", *Proc. Dixieme Colloque sur le Traitement du Signal et ses Applications*, Nice, France, May 1985.
34. H. J. Whitehouse and B. Boashash, "Delay-Doppler radar/sonar imaging", presented at EUSIPCO 88, Grenoble, France, 1988. See also J. M. Speiser, H. J. Whitehouse and J. C. Allen, "Wideband time-frequency distributions", . Chapter 6, *Time-Frequency Signal Analysis, Methods and Applications* (ed Boualem Boashash), Longman Cheshire/ Wiley, 1992.
35. T. C. M. Claasen and W. F. G. Mecklenbraucker, "The Wigner distribution - a tool for time-frequency analysis", Philips J. Research, Vol. 3, Part I, pp. 217-250, Part II, pp. 276-300, Part III, pp. 372-389, 1980.
36. B. Boashash, O. P. Kenny and H. J. Whitehouse, "Radar imaging using the Wigner-Ville distribution", in *Real-Time Signal Processing*, J. P. Letellier, Editor, Proc. SPIE Vol. 1154, pp. 282-294, 1989.
37. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
38. A. V. Jelalian, *Laser Radar Systems*, Artech House, Boston, MA, 1992.

Ideal

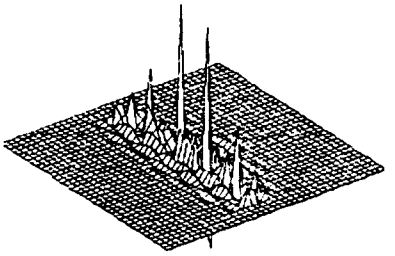


Point Spread Function

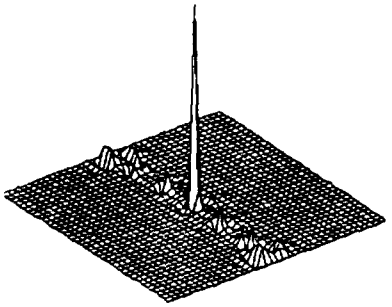
Degraded



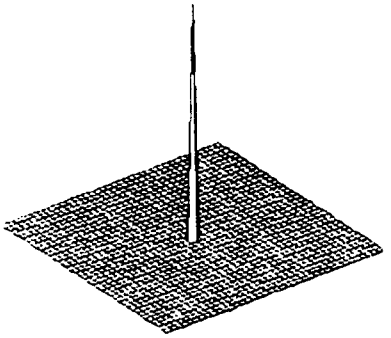
Inverse PSF - Pseudoinverse



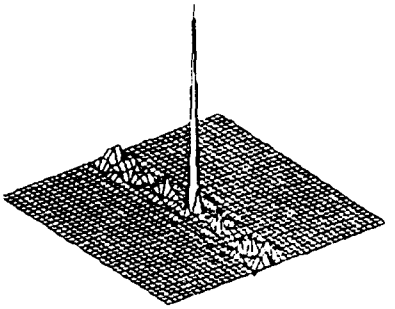
Inverse PSF - Wiener



Processed PSF - Pseudoinverse



Delta



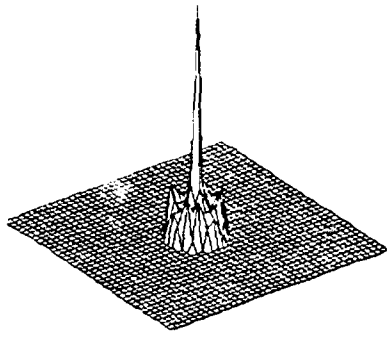
Processed PSF - Wiener

Processed - Pseudoinverse

Processed - Wiener

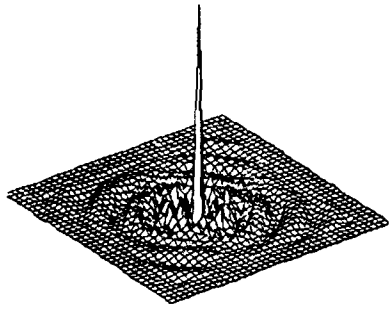
Figure 1. Point Spread Function Dominated by Astigmatism

Ideal

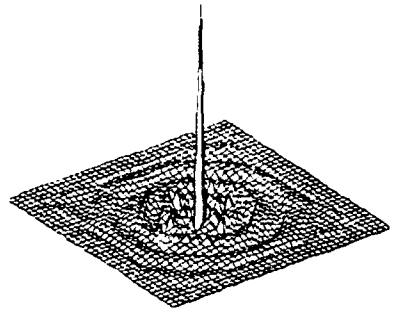


Point Spread Function

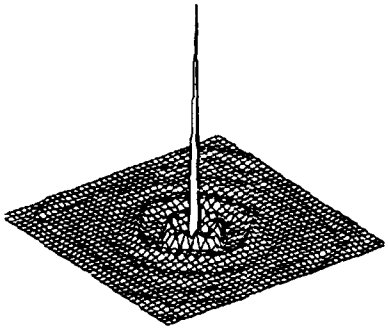
Degraded



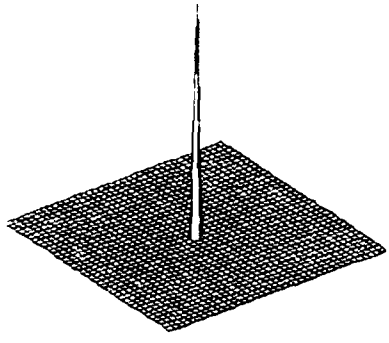
Inverse PSF - Pseudoinverse



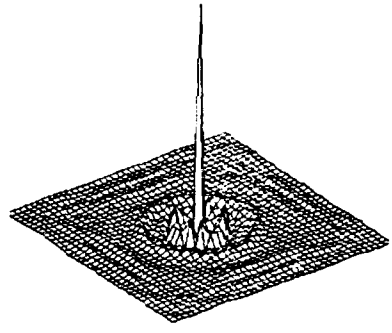
Inverse PSF - Wiener



Processed PSF - Pseudoinverse



Delta

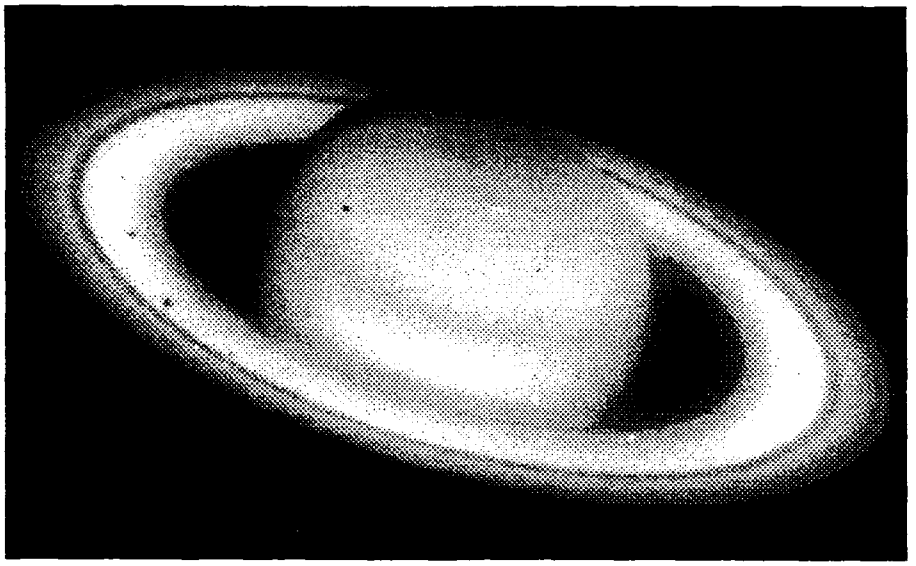


Processed PSF - Wiener

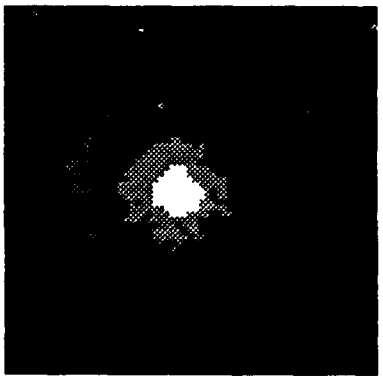
Processed - Pseudoinverse

Processed - Wiener

Figure 2. Point Spread Function Dominated by Defocus

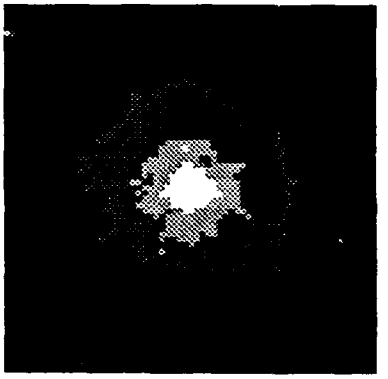


Unprocessed Saturn Image



Star #1 Image

Star #2 Image



Processed Saturn Image

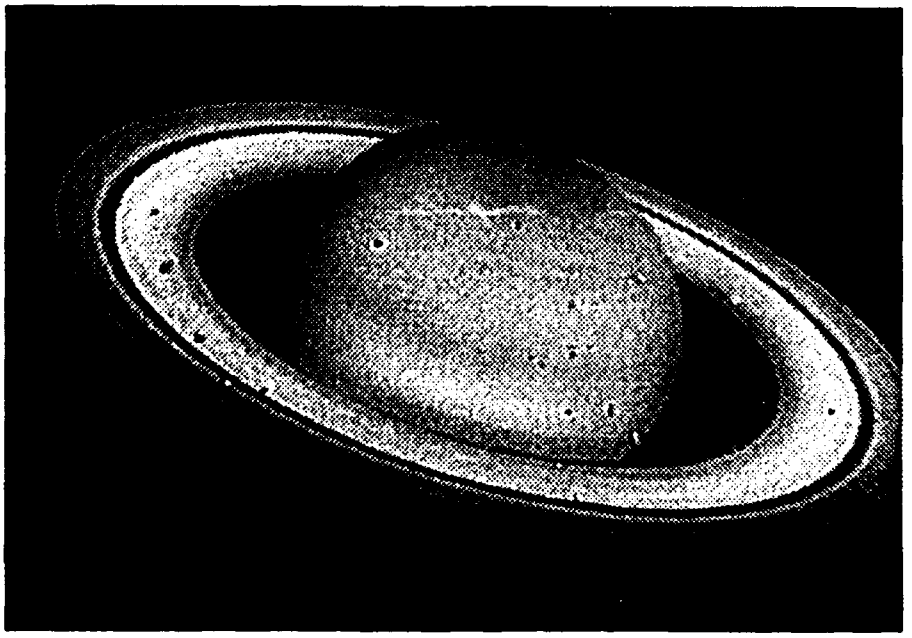
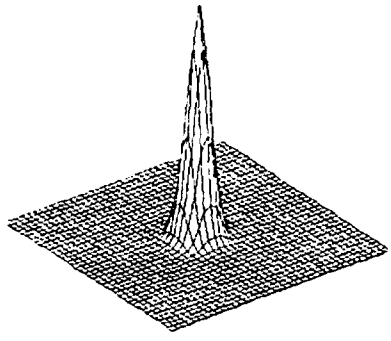


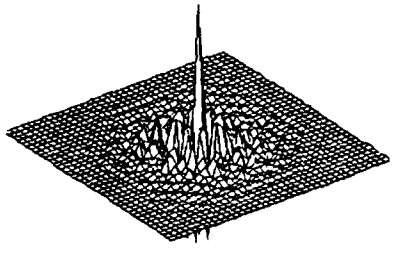
Figure 3. Hubble Space Telescope Imagery

Ideal

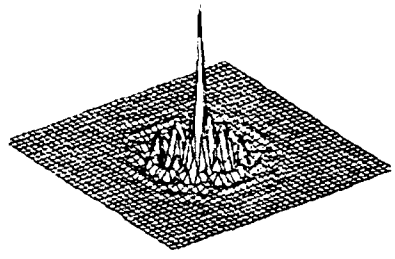


Point Spread Function

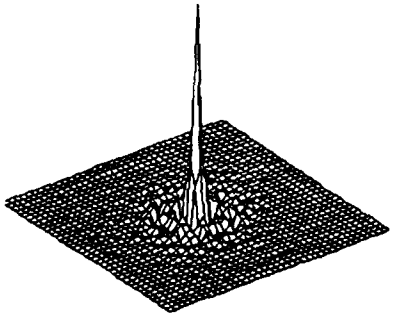
Degraded



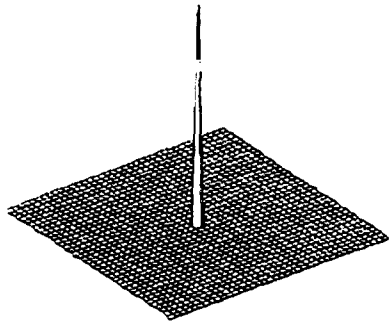
Inverse PSF - Pseudoinverse



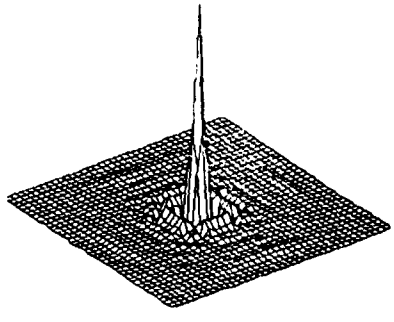
Inverse PSF - Wiener



Processed PSF - Pseudoinverse



Delta



Processed PSF - Wiener

Processed - Pseudoinverse

Processed - Wiener

Figure 4. Gaussian Point Spread Function

APPENDIX A - A CONSISTENT ESTIMATION FOR WVD RADAR IMAGING

Consider the radar deconvolution problem represented by Eq.(49). This equation assumes the expected value of the WVD of the received signal is known. This appendix discusses consistent estimation for this equation using a pulse-averaging scheme, but depends on substantial statistical assumptions regarding the random field of radar reflectors.

Under the assumptions outlined in Section 9, the received signal $\{r(t)\}$ is a stochastically filtered version of the transmitted signal $\{s(t)\}$ (Eq. (8) of [36]):

$$r(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D(x,y)\, s(t-x)\, exp[+i2\pi yt]\, dxdy. \tag{A1}$$

Here $D(x,y)$ denotes the random field determined by the reflectivity coefficient of the scatterers as a function of radial range x and radial Doppler y. For ease of exposition, we are assuming both $\{r(t)\}$ and $\{s(t)\}$ are analytic and have subsumed a deterministic phase term in $D(x,y)$. Then the WVD for the received signal has the form:

$$W_r(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D(x_1,y_1)\, D^*(x_2,y_2)\, s(t+\frac{\tau}{2}-x_1)\, s^*(t-\frac{\tau}{2}-x_2)$$
$$exp\left[+i2\pi y_1 (t+\frac{\tau}{2})\right] exp\left[-i2\pi y_2 (t-\frac{\tau}{2})\right] exp[-i2\pi f\tau]\, dx_1 dy_1 dx_2 dy_2 d\tau. \tag{A2}$$

The reflectivity field is assumed to be independent in range and Doppler:

$$E[D(x_1,y_1)\, D^*(x_2,y_2)] = \sigma(x_1,y_1)\,\delta(x_1-x_2)\,\delta(y_1-y_2). \tag{A3}$$

By taking the expectation of Eq. (A2) and using the covariance of $D(x,y)$ in Eq. (A3), it is straight-forward to obtain Eq.(49):

$$E[W_r(t,f)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma(x,y)\, W_s(t-x, f-y)\, dxdy = [W_s \star\star \sigma](t,f). \tag{A4}$$

It is one thing to write down an expectation and quite another to have a consistent estimator. The remainder of this appendix considers a pulse-averaging estimator for Eq. (49). Suppose N time-delayed pulses are transmitted so that the *nth* received waveform is given by

$$r_n(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} D_n(x,y)\, exp[+i2\pi yt]\, s_n(t-x)\, dxdy, \tag{A5}$$

where the *nth* pulse is $s_n(t) = s(t - t_n)$ and the associated random fields $D_n(x, y)$ satisfy:

(D-1) Each $D_n(x, y)$ is a zero-mean, circular Gaussian random field.

(D-2) The sequence $\{D_n(x, y)\}$ is independent and identically distributed with respect to n.

(D-3) Each $D_n(x, y)$ exhibits zero correlation with respect to x and y.

Assumptions (D-1), (D-2), and (D-3) imply that Eq.(A3) holds for each $D_n(x, y)$ for a single scattering function σ (that is, $\sigma = \sigma_n$) while the expectation between fields is zero: $E[D_n(x_1, y_1) D_m^*(x_2, y_2)] = 0$ for $m \neq n$. Consider averaging of time-delayed WVD of these received pulses:

$$W(N;t,f) = \frac{1}{N} \sum_{n=1}^{N} W_{r_n}(t + t_n, f). \tag{A6}$$

It is straight-forward to verify that the mean of this estimator is

$$E[W(N;t,f)] = [W_s \star\star \sigma](t,f). \tag{A7}$$

The covariance computation is more involved but it can be bounded:

$$\left| Cov[W(N;t_1,f_1), W(N;t_2,f_2)] \right| \leq \frac{4}{N} \|s\|_2^4 \times \|\sigma\|_1^2. \tag{A8}$$

Thus, *W(N;t,f)* is a consistent estimator for Eq. (49) as $N \to \infty$, provided the transmitted signal belongs to $L^2(\Re)$ and the scattering function σ belongs to $L^1(\Re \times \Re)$. While *W(N;t,f)* is consistent, note that assumption (D-2) begins to break down for large *N*. Indeed, while each $D_n(x, y)$ may still be Gaussian and independent with respect to n, it seems unlikely that the associated scattering functions could be identical as $N \to \infty$. It is conjectured, for a frequency independent reflectivity coefficient, that an alternative consistent estimator can be obtained by averaging the WVDs of an ensemble of frequency shifted pulses in a manner similar to Eq.(A6). This frequency averaged estimator should overcome the difficulty of accounting for the non-stationarity of the stochastic reflectivity field due to the motion in range of the scatterers during the observation interval of the ensemble.